

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:28:25 ; Search time 167 Seconds
(without alignments)
2004.080 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVIGLSPPQLCRPLP.....YEVFWSKSGWKITEGSLVAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4063	100.0	801	4	US-10-600-070-2
2	4051	99.7	801	4	US-10-600-070-127
3	4051	99.7	801	4	US-10-600-070-129
4	4051	99.7	801	4	US-10-600-070-131
5	4051	99.7	801	5	US-10-739-930-5771
6	2926	72.0	578	4	US-10-600-070-123
7	1775.5	43.7	760	4	US-10-600-070-125
8	1646	40.5	324	4	US-10-600-070-206
9	1332	32.8	525	4	US-10-600-070-122
10	1134	27.9	344	4	US-10-424-599-271849
11	973	23.9	364	4	US-10-425-115-266516
12	972	23.9	480	4	US-10-437-963-172416
13	856	21.1	416	4	US-10-425-115-242115
14	524.5	12.9	768	4	US-10-600-070-167
15	499	12.3	204	4	US-10-424-599-177901
16	498.5	12.3	789	4	US-10-600-070-194
17	497	12.2	798	4	US-10-600-070-164
18	497	12.2	798	4	US-10-600-070-165
19	482	11.9	709	4	US-10-437-963-172415
20	474.5	11.7	1157	4	US-10-437-963-155799
21	442	10.9	631	4	US-10-600-070-5
22	442	10.9	631	4	US-10-600-070-162
23	435.5	10.7	191	4	US-10-767-701-35633
24	434.5	10.7	624	4	US-10-600-070-116
25	431	10.6	819	4	US-10-600-070-171
26	431	10.6	819	4	US-10-600-070-173
27	429	10.6	714	4	US-10-600-070-169

28	429	10.6	714	4	US-10-600-070-170
29	425.5	10.5	652	4	US-10-600-070-115
30	423	10.4	673	4	US-10-600-070-192
31	413	10.2	836	4	US-10-600-070-190
32	409.5	10.1	205	4	US-10-767-701-41038
33	401	9.9	716	4	US-10-600-070-160
34	366	9.0	99	4	US-10-600-070-124
35	359	8.8	566	4	US-10-600-070-119
36	344.5	8.5	702	4	US-10-600-070-156
37	339	8.3	66	4	US-10-600-070-100
38	328.5	8.1	573	4	US-10-600-070-120
39	327	8.0	557	4	US-10-600-070-117
40	320	7.9	661	4	US-10-600-070-158
41	317.5	7.8	515	4	US-10-600-070-121
42	308.5	7.6	524	4	US-10-600-070-118
43	288	7.1	164	4	US-10-437-963-110670
44	263	6.5	66	4	US-10-600-070-99
45	257	6.3	66	4	US-10-600-070-98

ALIGNMENTS

RESULT 1

US-10-600-070-2

; Sequence 2, Application US/106000070

; Publication No. US20040139500A1

; GENERAL INFORMATION:

; APPLICANT: Osteryoung, Katherine W.

; APPLICANT: Vitha, Stanislav

; APPLICANT: Koksharova, Olga A.

; APPLICANT: Gao, Hong

; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of

; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of

; FILE REFERENCE: MSU-08153

; CURRENT APPLICATION NUMBER: US/10/600,070

; CURRENT FILING DATE: 2003-06-20

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 801

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; US-10-600-070-2

Query Match 100.0%; Score 4063; DB 4; Length 801;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEALSHVIGLSPPQLCRPLP	ATTKLRSHNTSTTICSASKWADRLSDFNFTSDSSSS	60
Db	1	MEALSHVIGLSPPQLCRPLP	ATTKLRSHNTSTTICSASKWADRLSDFNFTSDSSSS	60
Qy	61	FATATTATLVSPSIDRPERHVPI	PIDFYVQLGAQTHFLTDGIRAFARVSKPPQFG	120
Db	61	FATATTATLVSPSIDRPERHVPI	PIDFYVQLGAQTHFLTDGIRAFARVSKPPQFG	120
Qy	121	FSDALISRQIIQAACETLSNPR	SREYNEGLLDDDEATVITDVPMDKVPGLCVLQEG	180
Db	121	FSDALISRQIIQAACETLSNPR	SREYNEGLLDDDEATVITDVPMDKVPGLCVLQEG	180
Qy	181	GETEIVLRVGEALLKRLPKS	FKQDVVLVMAFLDVSROMALDPPDFITGYEFVEAL	240
Db	181	GETEIVLRVGEALLKRLPKS	FKQDVVLVMAFLDVSROMALDPPDFITGYEFVEAL	240
Qy	241	KLQEGGASSLAPDLRAQIDET	LEETIPRYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS	300
Db	241	KLQEGGASSLAPDLRAQIDET	LEETIPRYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS	300
Qy	301	VGGGGASALVGGLTREKFMNEA	FLRMTAAEQVDLIVATPSNIPAESFEVYEVVALVAQA	360
Db	301	VGGGGASALVGGLTREKFMNEA	FLRMTAAEQVDLIVATPSNIPAESFEVYEVVALVAQA	360

QY 361 FIGKKPHLLQADAKQFOOQOQAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
DB 361 FIGKKPHLLQADAKQFOOQOQAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
QY 421 CRMWGLGDSDSQYRNPAIYEFVLENSNRDNDLPGCLKLLETWLAGVVPFRDRTKDK 480
DB 421 CRMWGLGDSDSQYRNPAIYEFVLENSNRDNDLPGCLKLLETWLAGVVPFRDRTKDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAAAATWARI GAHVKASAMQALOKVPPSRYTD 540
DB 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAAAATWARI GAHVKASAMQALOKVPPSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNNGRDCGPGVFAEAVRPSNFETNDYAI RAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNNGRDCGPGVFAEAVRPSNFETNDYAI RAGVSESSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSSFORQKDMVSSMESDVATI 660
DB 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSSFORQKDMVSSMESDVATI 660
QY 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
DB 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSVLAS 801
DB 781 RYEVFWSKSGWKITEGSVLAS 801
RESULT 2
US-10-600-070-127
; Sequence 127, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hong
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-127
Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MEALSHVGIGLSPQLCRLPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
DB 1 MEALSHVGIGLSPQLCRLPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
QY 61 FATATTATLVSPPSIDRPERHVPIPIDEFYVLGAQTHETDGI RFAEARYSKPQFG 120
DB 61 FATATTATLVSPPSIDRPERHVPIPIDEFYVLGAQTHETDGI RFAEARYSKPQFG 120
QY 121 FSDALLISRQILQAA CETLSNPRSRREYN EGLDDEATVITDVPMDKVP GALCVLQEG 180
DB 121 FSDALLISRQILQAA CETLSNPRSRREYN EGLDDEATVITDVPMDKVP GALCVLQEG 180
QY 181 GETEIVLRVCEALLKERLPKSFQDQVVVLVMAALAFLDVSRDAMALDPDPFITGYEFVEAL 240

DB 181 GETEIVLRVCEALLKERLPKSFQDQVVVLVMAALAFLDVSRDAMALDPDPFITGYEFVEAL 240
QY 241 KLLQEBGASSLAPDLRAQIDETLEETIPRYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300
DB 241 KLLQEBGASSLAPDLRAQIDETLEETIPRYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNI PAESFEVYEVALALVAQA 360
DB 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNI PAESFEVYEVALALVAQA 360
QY 361 FIGKKPHLLQADAKQFOOQOQAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
DB 361 FIGKKPHLLQADAKQFOOQOQAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
QY 421 CRMWGLGDSDSQYRNPAIYEFVLENSNRDNDLPGCLKLLETWLAGVVPFRDRTKDK 480
DB 421 CRMWGLGDSDSQYRNPAIYEFVLENSNRDNDLPGCLKLLETWLAGVVPFRDRTKDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAAAATWARI GAHVKASAMQALOKVPPSRYTD 540
DB 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAAAATWARI GAHVKASAMQALOKVPPSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNNGRDCGPGVFAEAVRPSNFETNDYAI RAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNNGRDCGPGVFAEAVRPSNFETNDYAI RAGVSESSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSSFORQKDMVSSMESDVATI 660
DB 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSSFORQKDMVSSMESDVATI 660
QY 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
DB 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSVLAS 801
DB 781 RYEVFWSKSGWKITEGSVLAS 801
RESULT 3
US-10-600-070-129
; Sequence 129, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hong
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 129
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-129
Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MEALSHVGIGLSPQLCRLPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
DB 1 MEALSHVGIGLSPQLCRLPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60

QY 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDALLSRRQILQAACETLSNPRSREYNELGLDDEEATVITDVPWMDKVPFCALCVLQEG 180
DB 121 FSDALLSRRQILQAACETLSNPRSREYNELGLDDEEATVITDVPWMDKVPFCALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSKQDVVLVWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPKSKQDVVLVWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
QY 241 KLLQEGGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
DB 241 KLLQEGGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
QY 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVFPFRDRTDK 480
DB 421 CRMWLGDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVFPFRDRTDK 480
QY 481 KFKLGDYDDPWLVSYLEVERVEVQSGPLAAATWARI GAHVKASAMQALQKVPFSRYTD 540
DB 481 KFKLGDYDDPWLVSYLEVERVEVQSGPLAAATWARI GAHVKASAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSMESDVATI 660
DB 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSMESDVATI 660
QY 661 GSVRADDSALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPEVLIDGRMLKIWTDRAA 720
DB 661 GSVRADDSALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPEVLIDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYTLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYTLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLAS 801
DB 781 RYEVFWSKSGWKITEGSLAS 801

RESULT 4

US-10-600-070-131
; Sequence 131, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of Use
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 131
; LENGTH: 801
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

US-10-600-070-131

Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MEALSHVIGIGUSPQLCRLPPATTKLRSHNTSTTICSAKWADRLLDNFSTSSSSSS 60
DB 1 MEALSHVIGIGUSPQLCRLPPATTKLRSHNTSTTICSAKWADRLLDNFSTSSSSSS 60
QY 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDALLSRRQILQAACETLSNPRSREYNELGLDDEEATVITDVPWMDKVPFCALCVLQEG 180
DB 121 FSDALLSRRQILQAACETLSNPRSREYNELGLDDEEATVITDVPWMDKVPFCALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSKQDVVLVWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPKSKQDVVLVWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
QY 241 KLLQEGGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
DB 241 KLLQEGGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
QY 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVFPFRDRTDK 480
DB 421 CRMWLGDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVFPFRDRTDK 480
QY 481 KFKLGDYDDPWLVSYLEVERVEVQSGPLAAATWARI GAHVKASAMQALQKVPFSRYTD 540
DB 481 KFKLGDYDDPWLVSYLEVERVEVQSGPLAAATWARI GAHVKASAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSMESDVATI 660
DB 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSMESDVATI 660
QY 661 GSVRADDSALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPEVLIDGRMLKIWTDRAA 720
DB 661 GSVRADDSALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPEVLIDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYTLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYTLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLAS 801
DB 781 RYEVFWSKSGWKITEGSLAS 801

RESULT 5

US-10-739-930-5771
; Sequence 5771, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930

; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 5771
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C13643_1.p
US-10-739-930-5771

Query Match 99.7%; Score 4051; DB 5; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSFQLCRLPPATTKLRRSHNTTTCASAKWADRLSDPNFTSDSSSS 60
DB 1 MEALSHVIGLSFQLCRLPPATTKLRRSHNTTTCASAKWADRLSDPNFTSDSSSS 60

QY 61 FATATTATLVSIPPSIDRPERHVPIPIDFYQVLGAQTHFLTDGIRAFEARVSKPPQFG 120
DB 61 FATATTATLVSIPPSIDRPERHVPIPIDFYQVLGAQTHFLTDGIRAFEARVSKPPQFG 120

QY 121 FSDDALISRQIIQAACETLSNPRSRREYNEGLDDDEEATVITDVPWDKVPGALCVLQEG 180
DB 121 FSDDALISRQIIQAACETLSNPRSRREYNEGLDDDEEATVITDVPWDKVPGALCVLQEG 180

QY 181 GETEIVLRVGEALLKERLPSKQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPSKQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEAL 240

QY 241 KLLQERGASLAPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNILWS 300
DB 241 KLLQERGASLAPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNILWS 300

QY 301 VGGGGASALVGGITREKFNEAFLRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGGASALVGGITREKFNEAFLRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360

QY 361 FIGKPKHLLQADAKQFQOLQAQKVAIMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPKHLLQADAKQFQOLQAQKVAIMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420

QY 421 CRMWGLGSDSDSYRNPALVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOK 480
DB 421 CRMWGLGSDSDSYRNPALVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOK 480

QY 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATWARIAGAEHVKASAMQALQKVPFSRYTD 540
DB 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATWARIAGAEHVKASAMQALQKVPFSRYTD 540

QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFTAEAVRPSNFETNDYAIRAGVSSSV 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFTAEAVRPSNFETNDYAIRAGVSSSV 600

QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660
DB 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660

QY 661 GSVRADSEALPMDARTAEINISKWKIKSLAFGPDHRIEMLPEVLGDMKLIWTDRAA 720
DB 661 GSVRADSEALPMDARTAEINISKWKIKSLAFGPDHRIEMLPEVLGDMKLIWTDRAA 720

QY 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNADTVRTYTT 780
DB 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNADTVRTYTT 780

QY 781 RYEVFWKSGWKITEGSLVAS 801
DB 781 RYEVFWKSGWKITEGSLVAS 801

RESULT 6
US-10-600-070-123

; Sequence 123, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE OF INVENTION: Use
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 123
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-123

Query Match 72.0%; Score 2926; DB 4; Length 578;
Best Local Similarity 99.8%; Pred. No. 3.le-237;
Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 79 RPERHVPIPIDFYQVLGAQTHFLTDGIRAFEARVSKPPQFGFSDDALISRQIIQAACE 138
DB 1 RPERHVPIPIDFYQVLGAQTHFLTDGIRAFEARVSKPPQFGFSDDALISRQIIQAACE 60

QY 139 TSNPSRRREYNEGLDDDEEATVITDVPWDKVPGALCVLQEGGETEIVLRVGEALLKERL 198
DB 61 TSNPSRRREYNEGLDDDEEATVITDVPWDKVPGALCVLQEGGETEIVLRVGEALLKERL 120

QY 199 PKSFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEALKLLQEGASSLAPDLRAQ 258
DB 121 PKSFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEALKLLQEGASSLAPDLRAQ 180

QY 259 IDTLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNILWSVGGGGASALVGGITREKF 318
DB 181 IDTLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNILWSVGGGGASALVGGITREKF 240

QY 319 MNEAFURMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQAFIGKPKHLLQADAKQFQ 378
DB 241 MNEAFURMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQAFIGKPKHLLQADAKQFQ 300

QY 379 LQAQKVAIMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEMRMLGLDSEDSQYRNP 438
DB 301 LQAQKVAIMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEMRMLGLDSEDSQYRNP 360

QY 439 IVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOKKFKLGDDYDDPMVLSYLE 498
DB 361 IVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOKKFKLGDDYDDPMVLSYLE 420

QY 499 RVEVVGSPPLAAATWARIAGAEHVKASAMQALQKVPFSRYTDRNSAEPKDVQETVFSV 558
DB 421 RVEVVGSPPLAAATWARIAGAEHVKASAMQALQKVPFSRYTDRNSAEPKDVQETVFSV 480

QY 559 VGNVGRDGPVFTAEAVRPSNFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVK 618
DB 481 VGNVGRDGPVFTAEAVRPSNFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVK 540

QY 619 ILAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESD 656
DB 541 ILAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESD 578

RESULT 7
US-10-600-070-125
; Sequence 125, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.

Query Match	43.7%	Score 1775.5	DB 4	Length 760
Best Local Similarity	47.9%	Pred. No. 4.4e-140		
Matches	390	Conservative 119	Mismatches 213	Indels 93
Gaps	16			
QY	12	SPFQLCRLPPATTKLRESHNTSTTIC-SASKWADRLLSDFNF-----TSDSSSSSFATAT	65	
DB	14	APFAFSLPRPRPRPRPPPPHPSAACRAASRWAERLFADFHLLPTAAPSDPPSPAPAPAA	73	
QY	66	TTATVLSLPSIDPRERHVPIDIFYOVLGAQHTFLTDGIRRAFEARVSKPPQPGFSDDA	125	
DB	74	APSASPFVPLFPDAAERSLPLOVDYFKVLGAEPHFLGDGIRRAFEARIAKPPQVGYSTDA	133	
QY	126	LISRQILQAACETLSNPRREYNEGILLDDEEATVITDVPMDKVPGLCVLQBGGETE	185	
DB	134	LVGRQMLQIADHTLMQNSRTQYDRALSENREALTMDIAWD-----EAGEALA	184	
QY	186	VLRVGEALLKERLPKSPKQDWLVMAFLDVSDDAMALDPDPITGYEFVEERALKLQE	245	
DB	185	VLVTGEQLLDRPPKRFQDWLVAMALAYDLSDRDAASPPDVIGCEVLERALKLQE	244	
QY	246	EGASSLAPDLRAQIDETLEETPRYVLELLGLPGDDYAAKRLNGLSGVRNLIWSVCGGG	305	
DB	245	DGASNLAPDLLSQIDETLEETPRCVLELLSLPDIETHEHKRQBGLOQARNILMSVGRGG	304	
QY	306	ASALVGLGTRKFWNEAFLRMTAAEQVDLFFVATPSNIPABSFEYVEYVALALVAQAFICK	365	
DB	305	IATVGGGFSREAFWNEAFLRMTSIEQNDFFSKTPNSIPPEWFEIYNVALAHVAQAIIKSR	364	
QY	366	PHLLQADAKFOOLQQAQVWAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEKRWL	425	
DB	365	PQFTIMMADDLFEOLQKFNIGS----HYAYDN---EMDLALERAFCSLLVGDVSKCRMWL	416	
QY	426	GLOSEDSQYENPAIVEPVLNSN-RDDNDLPLGLCKLLETWLAGVFPFRDTRDKKFKPL	484	
DB	417	GIDNESPYPDKPILEVTNSSIENSEDLPLGLCKLLETWLI FEVFPFRSDTRGMQFRL	476	
QY	485	GDYVDDPMVLSYLERVEVWQSPLAATAWARTIGAE-----HYKASAMALQKVPF-SR	537	
DB	477	GDYVDDDEVLSYLERMEGGGASHALAAAATAKGAQATAALGTVKSNAIQAFNKFVPLIE	536	
QY	538	YTDNRSAEPPKDVQETVFSVDPVGNVGRGEPGVFIAEAVRPSENFTNDYAIRAGVSES	597	
DB	537	QLDRSAMENT-----KDG-PGGYL-----ENFQ-----	559	
QY	598	SVDETTVMGVADMLKEASVILAAQVAIGLISLFQSKYFLKSSSFQRKDMVSSMESD-	656	
DB	560	-----ENAPADHSRNAALKIISAGALFALLAVIGAKY-----LPRKRPLSAIRSEH	605	
QY	657	--VATIGSVRADDEAL-----PRMDARTENIVYSKWIKISLAGPDPHRIEMLEPVL	707	
DB	606	GSVAVANSVDSITDDPALDEDPVHPRMDAKLADIDVRKWQSIKSLAGPHSVASLQEV	665	
QY	708	DGRMLKIWTDRAAETAQGLVYDYTYLLKLKLSVDSVTVSADGTRALVEATLSEASCLSLVH	767	
DB	666	DGNMLKWITDRAAEIEHGHWFWEYTLSDVTIDISITSLDGRRTATEATIDEAGOLTDTVE	725	
QY	768	PENNATDVRTYTRYEVFWSK-SGWKITEGSLVAS	801	
DB	726	PRNDSYDTKYTRYEMAFSKLGGWKITEGAVLKS	760	

US-10-600-070-122

Query Match 32.8%; Score 1332; DB 4; Length 525;
Best Local Similarity 48.8%; Pred. No. 5.9e-103;
Matches 287; Conservative 88; Mismatches 137; Indels 76; Gaps 11;

QY 81 ERHVPIPIFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDDALISROILQAACETL 140
DB 3 ERSPLQVDYFVILGAEPHFLGDIRRAFEARVSKPPQFGSDDALISROILQAACETL 62

QY 141 SNPRSRREYNEGLDDEEATVITDVPWVKPGALCVLQEGGTEIVLRVGEALLKERLPK 200
DB 63 MNQNSRTQYDRALSENREALTWDAWK-----EAGALAVLVGTGEQLLDRPK 113

QY 201 SFQDVLVMAALFVSDAMALDPPDFITGVFEVEALKLLQEGCASLAPDLRAQID 260
DB 114 RFQDVLVMAALYVDSLSDMAASPPDVIGCCEVLERALKLLQEGCASLAPDLRLS 173

QY 261 ETLERITPRYVLELGLPGDDYAAKRLNGLSGVRNLSVGGGASALVGLTREKFN 320
DB 174 ETLERITPRYVLELGLPGDDYAAKRLNGLSGVRNLSVGGGASALVGLTREKFN 233

QY 321 EAPLMTAAEQVDLTVATPSNIPAESFEVEVALVAQAFIKKPHLLQADKQFOQLQ 380
DB 234 EAPLMTAAEQVDLTVATPSNIPAESFEVEVALVAQAFIKKPHLLQADKQFOQLQ 293

QY 381 QAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCECRMWGLDSEDSQYRNPAIV 440
DB 294 KFNIGS----HYAYDN----ENDLALERAFCSLLVGDVSKCRMWGLDSEDSQYRNPAIV 345

QY 441 EFVLENSN-RDNDLPLGLCKLLETWLAGVPRFRDTRKFKLGDYDDPWLSYLER 499
DB 346 EFVLENSN-RDNDLPLGLCKLLETWLAGVPRFRDTRKFKLGDYDDPWLSYLER 405

QY 500 VSVVQGSPLAAATWARIKAE-----HYKASAMQALQVFP-SRYTDRNSAEKDVQET 552
DB 406 MEGGASHLAAALAAKGAQATGALGVKSNIAQAFNKFVPLIEQLDRSAMENT----- 460

QY 553 VPSVDPVGNVGRDGPVGFIAEAVRPSNFETNDYAIRAGVSESSVDETTVMMSVADML 612
DB 461 -----KDG-PGYL-----ENFDQ-----ENAPAHDS 481

QY 613 KEASVKILAAGVAIGLISLFSOKYFLKSSSFQKMDKMSMESDVATI 660
DB 482 RNAALKIISAGALFALLAVIGAKY-----LPRKRPLSAIRSEHGSV 522

RESULT 10
US-10-424-599-271849
; Sequence 271849, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271849
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87499C.1.pap
US-10-424-599-271849

Query Match 27.9%; Score 1134; DB 4; Length 344;
Best Local Similarity 66.4%; Pred. No. 1.5e-86;
Matches 223; Conservative 48; Mismatches 53; Indels 12; Gaps 3;

QY 16 LKRLPPATT-----KLRRS--HNTSTTCSASKWADRLLSDFNTSDSSSSSFATATT 66
DB 12 LCTPHPTTHPTPKPKNKLRLSSLSRGAASLSATSQKWAERLIADFOFLGDAAS---TSTS 68

QY 67 TATLVLSPSIDRPERHVPIDFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDDAL 126
DB 69 TUSPSPVPRLDPPERIVSIPLDLYRILGAEPHFLGDIRRAFEARVSKPPQFGSDDAL 128

QY 127 ISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWVKPGALCVLQEGGTEIV 186
DB 129 ISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWVKPGALCVLQEGGTEIV 188

QY 187 LRVGEALLKERLPKSFQDVLVMAALFVSDAMALDPPDFITGVFEVEALKLLQEG 246
DB 189 LEITQGLLRERLPKTFQDVLVMAALFVSDAMALDPPDFITGVFEVEALKLLQEG 248

QY 247 GASSLAPDLRAQIDETLEETIPRYVLELGLPGDDYAAKRLNGLSGVRNLSVGGG 306
DB 249 GATSLAPDLRAQIDETLEETIPRYVLELGLPGDDYAAKRLNGLSGVRNLSVGGG 308

QY 307 SALVGLTTRKFMNEAFRLMTAAEQVDLTVATPSNI 342
DB 309 ATIAQFTREDFMNEAFRLMTAAEQVDLTVATPSNI 344

RESULT 11
US-10-425-115-266516
; Sequence 266516, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 266516
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_174665C.1.pap
US-10-425-115-266516

Query Match 23.9%; Score 973; DB 4; Length 364;
Best Local Similarity 54.8%; Pred. No. 5.9e-73;
Matches 198; Conservative 54; Mismatches 79; Indels 30; Gaps 4;

QY 12 SPQCLRLPPATTKLRRSINT-----STTICSASKWADRLLSDFNF-----TSDSS 57
DB 14 APFAFSLP-----LRRSHRPPPPGPGSTCAASRWADRLFADPHLLPAAADPPAAASSS 68

QY 58 SSSFATATTATVLSLPPSIDRERHVPIDFYVLGAQTHFLTDGIRRAFEARVSKPP 117
DB 69 SSPF-----VPIFPEADRALPLFVDYFKILGAEPHFLGDIRRAFEARVSKPP 117

QY 118 QGFGSDDALISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWVKPGALCVL 177
DB 118 QYGYSTEALVGRQMLQIAHDTLTNOSRTEYDRLSDEDRDALTDVMDKVGVLVCL 177

QY 178 QEGGTEIVLRVGEALLKERLPKSFQDVLVMAALFVSDAMALDPPDFITGVFEVE 237
DB 178 QEAGEAQLVATGEHLLQDRPKRFQDVLVMAALFVSDAMALDPPDFITGVFEVE 237

QY 238 EALKLQEGCASLAPDLRAQIDETLEETIPRYVLELGLPGDDYAAKRLNGLSGVRN 297
DB 238 RALKLQEGCASLAPDLRAQIDETLEETIPRYVLELGLPGDDYAAKRLNGLSGVRN 297

RESULT 14

Search completed: December 4, 2005, 12:42:30
Job time : 170 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:29:10 ; Search time 11 Seconds
(without alignments)
348.678 Million cell updates/sec

Title: US-10-600-070B-2
Perfect score: 4063
Sequence: 1 MEALSHVIGLSFQICRLP.....YEVFWSKSGWKITEGVLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 26661 seqs, 4788334 residues

Total number of hits satisfying chosen parameters: 26661

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.New.*
1: /cgn2_6/protdata/1/pubpaa/US09_NEW_PUB pep.*
2: /cgn2_6/protdata/1/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/protdata/1/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/protdata/1/pubpaa/US08_NEW_PUB pep.*
5: /cgn2_6/protdata/1/pubpaa/PCT_NEW_PUB pep.*
6: /cgn2_6/protdata/1/pubpaa/US10_NEW_PUB pep.*
7: /cgn2_6/protdata/1/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/protdata/1/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	116	2.9	1448	6	US-10-485-517-212
2	111.5	2.7	3704	6	US-10-513-786-1
3	109.5	2.7	661	6	US-10-467-657-1266
4	107	2.6	908	6	US-10-467-657-1070
5	105.5	2.6	817	6	US-10-793-626-50
6	105.5	2.6	817	6	US-10-793-626-1528
7	104.5	2.6	402	6	US-10-858-730-30
8	104	2.6	627	6	US-10-467-657-5432
9	103.5	2.5	715	6	US-10-467-657-70
10	103.5	2.5	715	6	US-10-467-657-3672
11	103.5	2.5	2004	6	US-10-467-657-84
12	103.5	2.5	2004	6	US-10-467-657-6322
13	103	2.5	565	7	US-11-074-176-98
14	101	2.5	403	6	US-10-858-730-27
15	101	2.5	711	6	US-10-467-657-2966
16	100.5	2.5	1663	6	US-10-982-545-6
17	99	2.4	943	6	US-10-467-657-5508
18	98	2.4	611	7	US-11-082-389-436
19	97.5	2.4	505	6	US-10-467-657-5148
20	97.5	2.4	791	6	US-10-467-657-5014
21	97.5	2.4	1192	6	US-10-858-730-72
22	96.5	2.4	724	6	US-10-793-626-3082
23	96.5	2.4	1152	7	US-11-055-822-308
24	95.5	2.4	566	6	US-10-467-657-3302
25	94.5	2.3	1020	6	US-10-513-786-4

26	94.5	2.3	1458	7	US-11-096-274-2	Sequence 2, Appli
27	94.5	2.3	1900	6	US-10-513-786-3	Sequence 3, Appli
28	94.5	2.3	4384	6	US-10-821-234-1120	Sequence 1120, Ap
29	94	2.3	423	7	US-11-110-851-2	Sequence 2, Appli
30	94	2.3	745	6	US-10-858-730-224	Sequence 224, App
31	94	2.3	745	7	US-11-055-822-226	Sequence 226, App
32	94	2.3	745	7	US-11-055-822-714	Sequence 714, App
33	94	2.3	745	7	US-11-055-822-756	Sequence 756, App
34	94	2.3	1217	7	US-11-074-176-252	Sequence 252, App
35	93	2.3	617	6	US-10-982-545-2	Sequence 2, Appli
36	92.5	2.3	403	6	US-10-858-730-28	Sequence 28, Appli
37	92	2.3	481	6	US-10-467-657-3916	Sequence 3916, Ap
38	91.5	2.3	406	6	US-10-467-657-5504	Sequence 5504, Ap
39	91.5	2.3	588	6	US-10-793-626-3328	Sequence 3328, Ap
40	91.5	2.3	1279	6	US-10-793-626-3188	Sequence 3188, Ap
41	91	2.2	817	6	US-10-793-626-2948	Sequence 2948, Ap
42	90.5	2.2	426	6	US-10-467-657-1548	Sequence 1548, Ap
43	90.5	2.2	456	6	US-10-467-657-5786	Sequence 5786, Ap
44	90.5	2.2	859	6	US-10-467-657-6084	Sequence 6084, Ap
45	90	2.2	296	6	US-10-467-657-5024	Sequence 5024, Ap

ALIGNMENTS

RESULT 1
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 02000349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-485-517-212

Query Match		2.9%	Score 116;	DB 6;	Length 1448;
Best Local Similarity		19.4%	Pred. No. 0.52;		
Matches 134;		Conservative 103;	Mismatches 234;	Indels 220;	Gaps 33;
Qy	151	EGLLDDEATVITDVPMDKVPGLCVLQGGETHIVRVGSAALKERLPKSKQDVLMV	210		
Db	415	QGITAD---TTIKDVAKDEL-----ATKANEQKALIAQTADATTEKEQANQOVDAQLTQ	466		
Qy	211	ALAFDVSRRAMALDPDFITGYEFVEEALKLLO-----EEGASSLAPDLRAQIDETLE	264		
Db	467	GNQNIIE---NAQSID--DVNTAKDNAIQAIDPIQASTDVTKNARAELLTEMONKITEILN	521		
Qy	265	EITPRVYLELLGLPLGDDYAAKRLNLSGVNRILWSVGGGASALVGLTKEKFNFAFL	324		
Db	522	N-----NETTNEEKGNIDIGPVRAAYEEGLNNI-----NAATTGDVTTAK-----	561		
Qy	325	RMTAEEQVDLVFATPSNIPAESFEVEVALVAQAFIGKPKPHLLQDADKQFQQLQQAQV	384		
Db	562	-DTAVQVKVQQLHANPVKKPAGKKE-----LDQAAADKKTQIEQTPNASQOEINDAK-	611		
Qy	385	MAMEIFAMLYDTRNNWEIDFGLERGLCALLIGKVDRCMMLGLDSEDSQYRNPAIVEFL	444		
Db	612	--QEVDTLNOAKTN-----VDQ-----SSTNEYVDNAVKE---	640		

Qy	306	ASALV----	GGLTREKFMN-----	EAPLRMTAAEQVDFLV-----	ATPSNTPASBSF	347
Db	268	AFALTWLEEGSCANESYYNL	IPTSLGGTHEAGLKQAFVAVNNF	INLHLLPRGVKVQSD	327	
Qy	348	EWV-EVALALVAQAFIGKPHLL	QDADKQFOLOQAKVWAME----	I PAMLYDTRNNW----	400	
Db	328	DVFGKTAFLVSARVL-----	DPOFGQTGDKLTNRDALKLVA	AVSGDPLLEMLN	376	
Qy	401	-EIDPG-----	LER-----	GLCALLIGKVDRCRMWLGDS	SDSOVR	435
Db	377	QNVDFGKKIAELAIHQAOAR	ISRVKIKKKSGSV-AILPGKLT	QD-----	ESSEDIREN	429
Qy	436	NPAIVEFVLENSNRDDNDL	PLCLCKLLETW-----	LAGVVFPFRDTKDKKFKLG	DIY	488
Db	430	ELFLVE-----	GDSAGSAKLARDKATQAIL	PLRGKVLNSFEVHPDQ	FGNAEIH	479
Qy	489	D-----	DPMVL SylERVEVUQSGS	PLAAAAATMARIGAEHVKAS	AMQALQKVFP	541
Db	480	DISVAVGVDPHAINDH	PLSLRGLYGKIAILSD-ADVDG	SHIQVLLTLFYRHFPKL	VADG	538
Qy	542	N--SAEPKQVQETVESVDV	PVGNVCRDCEPGVFIAEAVR	PSENPENETNDYATRAGV	SESSV	599
Db	539	HIYIAQP-----	PLFRVD-----	VNAQGS-----	KPARKF-----	567
Qy	600	DETTVMESADMLKEASVK	LAAAGVAIGLISLFSOKYFLK	SSSSFORXDMVSSB	SDVAT	659
Db	568	DONELD-GILERLOKEGV	KETAYSI-----	SRFKGLGEMNPDQLKD-	TTTHPPDTR	616
Qy	660	IGSVRADDSALPRMDART	AEINSVKWKOKIKSLAFGPD	PHRIEMLPEVL	DGRMLKI	719
Db	617	LLQVOIPEG-----	ADDETRDI	FVKLMGKEAAA-----	RRAWMERE	653
Qy	720	AETAQLGL	727			
Db	654	GDTAQLDI	661			

RESULT 4

```

US-10-467-657-1070
; Sequence 1070, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASTIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AL
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1070
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1070

```

Query Match 2.6%; Score 107; DB 6; Length 908;

Query Match	2.8%;	Score 107;	DB 6;	
Best Local Similarity	19.3%;	Pred. No. 1.2;		
Matches 154;	Conservative 108;	Mismatches 250;	Indels 286;	Gaps 37;

```

143 QY PRSRREYNGLLDDEEAT-----VITDPMDKVFGALCVLQEGE--TEIVLVRGE----- 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 Db PDLRPOAEALPDLVRLTGWPLVIGQEAADVITGLA--KQGAHGLRVLVSTGDKDMA 119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 QY ALLKERLPKFKQDVVLVNALAFIDVS-----RDAMAL--DPPDFTTGFVEFV 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
120 Db QLVDERV-----TLVNTMGETLIDIEGRKAKFGRVPOIRDYLYALIGDKVDNVPVGEKC 173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Query Match

2.6%; Score 105.5; DB 6; Length 817;

QY	237	--	BEAKULLQE-----EGASSIAPDLRAQIDETLEETIPRYVL--BLLGLPLGDDVAARKL	288
Db	174	GPKTAVKWL	EAYGSLQGVMEHAPETKGVGENLQAALPOLPLSYDLVLTIKTDVDLHTLS	233
QY	289	NGLSVR-----	NILWSVGGGASALVGL	313
Db	234	DGIESLRRTAPKWAQLAVDFKRWGFRTWLKEASNMNTGSTDDLFGSDSIGQAALNAEI	293	
QY	314	TREKFMNEAFLRMTAAEOVDLFVAT-----	PSNIIPASSFEVY	350
Db	294	PSEKGAEXA-----TAPEKLDYQAVTTEAQFAALLDKLAKADTIGIDTETTSLDANNAELV	349	
QY	351	EVALALVA-----	QAFIGK-KPHLLQADQKQFQLOQAOKWAME	388
Db	350	GISIAFQAGEAVYIPVGHSLTAAPEQLDLQDVLGRLKPHLENPALKKIQ-----	399	
QY	389	IPAMLYDTR-----	NNWEIDFGLERGLCALLIGK-VDECR-MWLGLD-----	428
Db	400	--NLYQDQHVFPANYGIALNNIAGMALASYITESHGLGHDELSGRWLTLETITYESLCG	457	
QY	429	--SEDSQYRNPAI---VEFVLENSNRDDNDLPGCLKLETLWAGVVPFRDTRKDKFKF	483	
Db	458	KGAKOISFADVAICQATEYAAQDDA-----FAURLEARLRA-----QMDKQLE	501	
QY	484	LGDIYDDPMLWLSYLE-RVEVYVQGSPLAAAAWTARITGAHVXASAMQALQKVPFSPRYTORN	542	
Db	502	MYEKMELPVAQVLFEMECNGVQIDRAELARQSAELGAELMKLE-----QEAVAAGQFPN	556	
QY	543	SAEPKDVQVETVFSVDPVGNVGRDGPGVFAEAVRPSENPETNDYAIRAGVSESSVDET	602	
Db	557	LNSPKQLQEIIF--DKMG-----IPTKGLKK--TAKGGISTN-----	589	
QY	603	TVEMSVADMKLEASVKILAAQVAILGLISLPQKYP-LKSSSSSFQKMDVSSMESDVAT--	659	
Db	590	-----EAVLEQAPDYPFLPKIILQNRSLAKLKSTYTDKLPPEMISPRDNRVHTTY	638	
QY	660	-----IGSVRADDSALPRMDARTAEINIVSKWQIKSLAFPGPDHRIEMLPEVLDCRMKI	714	
Db	639	AQAVAITGRLASNNPNLQNIPIRAE-----	GRRV--	668
QY	715	WTDRAAETAQGLVY---DYTLKLKLSVDSTVTSADGTRALVEATLEESACLSDLVH---	767	
Db	669	--RRAFTAPPGSVIVSADYSQIEL---RIMAHLSGDKTLIAAFQS-----GEDVHRRTA	717	
QY	768	-----PENNATDVRY	778	
Db	718	AEEVFGTAPENVPNSQRRY	735	

RESULT 5

```

US-10-793-626-50
; Sequence 50, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 817
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: amino acid sequence
US-10-793-626-50

```

Query Match 2.6%; Score 105.5; DB 6; Length 817;

Best Local Similarity 18.9%; Pred. No. 1.3; Mismatches 248; Indels 251; Gaps 37;
Matches 141; Conservative 108; Mismatches 248; Indels 251; Gaps 37;
Query Match 2.6%; Score 105.5; DB 6; Length 817;
Best Local Similarity 18.9%; Pred. No. 1.3; Mismatches 248; Indels 251; Gaps 37;
Matches 141; Conservative 108; Mismatches 248; Indels 251; Gaps 37;
US-10-793-626-1528

124 DALISRRQILQACETLSNPRSRREYNEGLDDDEATVITDPWPKVPGALCVLQGGT 183
180 DPVGRDKETIRVIEVL-----SRTKNNPVLIGE-----PGV-----GKT 215
184 EIVLRVGEALLKERLPKSPKQDVVLVMAFLDVSRDAMALDPPDFTITGYEF---VEEAL 240
216 AIAEGLAQAIKNEVPETLK-----DKRVMSLDMGTIVAGTKYRGEFEERL 261
241 KLLQEE-----GASSLAPDLRAQIDETLEITPRYVLELGLPLGDD 282
262 KVMEEIHOAGNVILFIDELHTLVGAGGAGDA--SNILKPALARGELQCIGATTLDE 319
283 Y-----AAKE-----LNGLSGVNRLMSVGGGASALVGLTRE 316
320 YRKNIEKDAALERFQPIQVDEPTVEDTIELKGLDRDYEAHHRINISDEALEAAKLSD 379
317 KFMNEAFLRMTAAEQVDLFVA-----TPSNI PAESFEVVEVA-----LALVAQAFIGK 364
380 RYVSDRFLPKAIDLIDEASSKVRKLSHTTPSNLKEIEQIDKVKNEKDAVHAQEF--- 436
365 KPHELLQADKQFOLOQAKVAMEIPAMLYDTNNWE-----IDFGL-ERGLCALLIGKV 418
437 -ENANLRDKQSKLEKQYE-----DAKNEWKNAQGGDLTALSSENTAEVITAG-- 482
419 DECRMWGL-----DSEDSQYENPAIVFVLNSNRDNDLPGCLKLETWLAGVFP 472
483 -----WTGIPLT KINETESDRLLN---LEDTLHKRVIGQNDVANSISKAVRRARAGL--- 531
473 RFRDTDKKFKLGDY-YDDPMVLSYLERVEVVGSGPLAAAATWARI G-AEHVKASAMQAL 530
532 -----KDKRPFGSFIPLGTGVGKTELARALAESMFGEDDAMIRVDMSEFMEKHAVSRL 586
531 QKVFP-----SRYTDRNSAEPKDV-----QETVFSV-----DPVGN 562
587 VGAPPYGVHDDGGQLTEKVRKPYSVILFDETEKAHPDVFNILLQVLDGHLTDTKGRT 646
563 VGRDGPVGFIAEAVRPSNFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVKILAA 622
647 V--DFRNTVIMTSNVGAQELQORFAGGASEGS-DYETVRKT---MKKE----- 692
623 GVAIGLISLFSQKYFLKSSSSFORKDMVSSMESDVATIGSVRADDSALPRMDARTAE NI 682
693 -----LKNSEF---RPEFLNRVD-DIIVFHKLT KDELKEIYTM-----M 726
683 VSKWQIKSLAFGPDHRIEMLPEVLDRMLKIWTDRAAE-TAQLGLVYDYTLKLSVDSV 741
727 VNKL-----THRLS-----EQNINIVVTDKAKEIABEGYDPEY----- 760
742 TVSADGTRALVEA---TLESACLSDLV 766
761 -----GARPLIRAIQKTVEDN--LSELI 781

RESULT 6

US-10-793-626-1528
; Sequence 1528, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUB480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1528
; LENGTH: 817
; TYPE: PR

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1528

Query Match 2.6%; Score 105.5; DB 6; Length 817;
Best Local Similarity 18.9%; Pred. No. 1.3; Mismatches 248; Indels 251; Gaps 37;
Matches 141; Conservative 108; Mismatches 248; Indels 251; Gaps 37;
US-10-793-626-1528

124 DALISRRQILQACETLSNPRSRREYNEGLDDDEATVITDPWPKVPGALCVLQGGT 183
180 DPVGRDKETIRVIEVL-----SRTKNNPVLIGE-----PGV-----GKT 215
184 EIVLRVGEALLKERLPKSPKQDVVLVMAFLDVSRDAMALDPPDFTITGYEF---VEEAL 240
216 AIAEGLAQAIKNEVPETLK-----DKRVMSLDMGTIVAGTKYRGEFEERL 261
241 KLLQEE-----GASSLAPDLRAQIDETLEITPRYVLELGLPLGDD 282
262 KVMEEIHOAGNVILFIDELHTLVGAGGAGDA--SNILKPALARGELQCIGATTLDE 319
283 Y-----AAKE-----LNGLSGVNRLMSVGGGASALVGLTRE 316
320 YRKNIEKDAALERFQPIQVDEPTVEDTIELKGLDRDYEAHHRINISDEALEAAKLSD 379
317 KFMNEAFLRMTAAEQVDLFVA-----TPSNI PAESFEVVEVA-----LALVAQAFIGK 364
380 RYVSDRFLPKAIDLIDEASSKVRKLSHTTPSNLKEIEQIDKVKNEKDAVHAQEF--- 436
365 KPHELLQADKQFOLOQAKVAMEIPAMLYDTNNWE-----IDFGL-ERGLCALLIGKV 418
437 -ENANLRDKQSKLEKQYE-----DAKNEWKNAQGGDLTALSSENTAEVITAG-- 482
419 DECRMWGL-----DSEDSQYENPAIVFVLNSNRDNDLPGCLKLETWLAGVFP 472
483 -----WTGIPLT KINETESDRLLN---LEDTLHKRVIGQNDVANSISKAVRRARAGL--- 531
473 RFRDTDKKFKLGDY-YDDPMVLSYLERVEVVGSGPLAAAATWARI G-AEHVKASAMQAL 530
532 -----KDKRPFGSFIPLGTGVGKTELARALAESMFGEDDAMIRVDMSEFMEKHAVSRL 586
531 QKVFP-----SRYTDRNSAEPKDV-----QETVFSV-----DPVGN 562
587 VGAPPYGVHDDGGQLTEKVRKPYSVILFDETEKAHPDVFNILLQVLDGHLTDTKGRT 646
563 VGRDGPVGFIAEAVRPSNFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVKILAA 622
647 V--DFRNTVIMTSNVGAQELQORFAGGASEGS-DYETVRKT---MKKE----- 692
623 GVAIGLISLFSQKYFLKSSSSFORKDMVSSMESDVATIGSVRADDSALPRMDARTAE NI 682
693 -----LKNSEF---RPEFLNRVD-DIIVFHKLT KDELKEIYTM-----M 726
683 VSKWQIKSLAFGPDHRIEMLPEVLDRMLKIWTDRAAE-TAQLGLVYDYTLKLSVDSV 741
727 VNKL-----THRLS-----EQNINIVVTDKAKEIABEGYDPEY----- 760
742 TVSADGTRALVEA---TLESACLSDLV 766
761 -----GARPLIRAIQKTVEDN--LSELI 781

RESULT 7

US-10-858-730-30
; Sequence 30, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.

```

; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 402
; TYPE: PRF
; ORGANISM: Streptomyces coelicolor
US-10-858-730-30

Query Match          2.6%; Score 104.5; DB 6; Length 402;
Best Local Similarity 21.0%; Pred. No. 0.53; Indels 111; Gaps 14;
Matches              70; Conservative 50; Mismatches 103;

QY   94 LGAQTFLTDGIRRAFEARV-----SKPQGFGSDALISRRILQAACETLL 140
    :|:::||:|||||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   96 IGAQSPDIAQGVDTAYENRVEGEDELDROAGDGQGLMGFYASDETPTLMPLPVLAHRL 155

QY   141 SNPRSREYNELGLDDDEATVITDPWPKVPKALCVLQEGETEIVLRV--GEALLKERLP 199
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   156 SKRLSEVRKN-----GTIPYLRPDKTQVTIYEYDC-----186

QY   200 KSFKQDVVLWMALAFLDVSRDALMDPPDITIGVFVEERALKLLOEGASSLAPDLRAQI 259
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   187 KAVRLDTVVSSQHRSDI--DLSSLAPDI---KEFVVE-----PELKALL 227

QY   260 DETLEEITPRVLEL-----LCPLPGDDYAARKNLGLSGVRNLMSVG-----CGGA- 306
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   228 EDGIKDITYENRYLVNPTRFRFGPEGWD-----AGLTGRKIITIYGGMARHGGAFF 280

QY   307 -----SALVGGLTREKFPMNEAFLRMTAAEQVDLFVAT--PSNIP 343
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   281 SGKDPKVDRSAAYAMRWVAQNVAAGLAARCEVOVAY-AIGKAEPVGLFVFETGTAKVD 339

QY   344 AESF-----EVYEVALVAQAFIGKKPHLLQDA 372
    |::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   340 TEKTEKAIDEVFDLRPAAITRALDLLRFPIYAOTA 373

```

```

RESULT 8
US-10-467-657-5432
; Sequence 5432, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5432
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae

```

US-10-467-657-5432

Query Match 2.6%; Score 104; DB 6; Length 627;
 Best Local Similarity 19.7%; Pred. No. 1.2;
 Matches 124; Conservative 87; Mismatches 208; Indels 210; Gaps 33

Qy	241	KLQEEGASSIAPDLRAQIDETLEBEITPRVYVLELLGLPLGDDYAAKRLNGLSGVNIIWS	300
Db	10	KVIENAEAGARTPSIIAYLDG--GEI-----LVGAP-----AKR-QAVTNAKNTIYA	53
Qy	301	VGGGGASALVGLLTREKFMNEAFLR-----MTAAEQVDLFV-----ATPSNIPAES	346
Db	54	-----AKRLIG-----HKPEDKEVQRDIETSEMPFEIHKADNGDAWKAQKELSPPOISAEV	104
Qy	347	FEVYEVALALVAQAFIGKCKPHLLQADKQFOOLQQAQKVMAMEIPAMLYDTRNNWEIDFGL	406
Db	105	LKKMKEA-----AEAYLGEK-----VTEAVITVPAYFNDSQOATKDAGR	144
Qy	407	ERGL-CALLIGKVDCEKRWJLGLSEDSQYRNPALVEF-----VLNSNRDNDNDLUPG	457
Db	145	IAGLDVKRIINEPTAAALAFGMDKGNDRKRIAYVDLGGGTFDISIIEIANLDGDKQPEV	204
Qy	458	LCKLLETWLAGVVF-PRFRDTOKKFKLG--DYDDPMVLSYLE-----RVENVQSP	507
Db	205	LATNGDTFLGGEDFDQRLDIYIIDEFKKEQIGDILQDVMALQRLKEAAEAKAKIELSSGQQ	264
Qy	508	L---AAAATWARI GAEHVKASAMQALQKVPFSRYTDRNSAEPKDVQETVFSVDPVGNVVG	564
Db	265	TEINLPYITMDATGPKHL-----AMKITRAKFESLVED-----	297
Qy	565	RDGEPGFVIAEAVRPSNFETNDYAIR-AGVSESSVDET-----TVEMSVADMLK	613
Db	298	-----LIARSIEPCRT-----ALKADAGLSTGDIIDVILVGGQSRMPKVQEAVIDPFG	344
Qy	614	EASVK-----ILAAQVAIGLILFSQKYLKSSSSFORKDM-----VSSMESDVATTGSV	663
Db	345	KEPRKDVNPDEAVAVGAAI-----QGEVLSGG-----RSDVLLILDVTPISLGIETMGVV	393
Qy	664	RADDSEALPRMDARTAEV--IVSKWQIKSLAFCGPDHRIEMLPVLIDGRMLKIWTDRAAE	721
Db	394	-----MTKLIQKNTTIPTKASQVFFSTA--EDNQSAVTHIVLQGE-----RERASA	436
Qy	722	TAQLG-----LVYDYT--LILKLSV-----DSVTVVSADGTRAL---	751
Db	437	NKSLGQFNVLGDIAPAPRGMPOIEVTFDIDANGILHVSADKDGKYGKAANIITQSSGLSEE	496
Qy	752	-----VEATLEESACLSDLVHPENNA	772
Db	497	EIRRMVKDAAEAAFEEDKKI TEIVASRNOA	525

```

RESULT 9
US-10-467-657-70
; Sequence 70, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 70
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae

```


Db 712 KALREHAESIKKRLAESIGGLAERVDVAASVETAPDKAQMLLSQRVGDFDRTGKITLV 771
QY 190 GEALLKERLPKSPKQDVVVLWALAFLOVSRDAMALDDPDPFITGYEFVEEALK----- 241
Db 772 AENLTPEP-----AVMAAHELGHGFAAD-----GPAKYREELERADGNGLI 814
QY 242 -----LLQE--EGASSLAPDLR-AQIDETLEETPRVYVLELLGLPLGDDVAAKRLNGLSG 293
Db 815 RRIADAVQEGREGTGDAASVRPAAVEAVAL-----YAAQRTGGWAG 858
QY 294 VRNLTW-SVGGG-----GASALVGGITR---EKFMNEAFLRMTAAEQVDLFPVATPSNIP 343
Db 859 IENRYGVKNGLRGKIAGVLARIGALLRRVLQRLAGKAGGAMSDADVAFMLADLHGNGVE 918
QY 344 AESFEVY--EVALALVAQAFTGKPKPHLLQDADKQFOOLQOAKVMAMEIP--AMLYDTRNN 399
Db 919 GARDAPMGNGHRAVMFAAEADGAERSKESLEKRLRAETIRISGREVPPEGNLRREYKRN 978
QY 400 WEIDFGLERGLCALLIGKVDECRMWLGL-----DSEDSQY-RNPAIVEFVLENS-- 447
Db 979 -ALEYG--KSLRGPVYNKDTGREISLGRSGITELRHLDYKDAEHLQSIATAIPQIIENAVY 1035
QY 448 -----NRD--DNDDLPGLCKLLETWLAGVVF-----PRFRDTKDKKFK 483
Db 1036 IDTLPNEDLAKNGDIQG---YEVYVSGLVNGGADYTVRAAVAVSRNGNRYYYDHKLTKIE 1091
QY 484 LGDYDDPMVLSYLERVEV---QGSPL-----AAAAWARIAGAEHVK 523
Db 1092 KGN-----LLSLDRVSTTGASESKSPSGIDDKRLQLQDKDAGKGGIADPDTBAVR 1145
QY 524 ----ASAMQALQVFPSSRYTDRNSAEPKQVETVFSVDPVGNVNGRGEPCGVF--TAEAV 577
Db 1146 FSRAANIEAIGRITGKKSLDRNALKDRWDASKGIQLQFLGRRQIEDIYGGVLDGLKEYG 1205
QY 578 RPSENFTNDYATRAGVSESSVDTEVMSVADMLKEASVKILAAGVAIGLISLFSQKYF 637
Db 1310 KAAYALEKAQE-----KHGREILADEADMRLRLFYADSEAKRA-----LRA 1352
QY 737 SVDSVTVS---ADGTRALVEA 754
Db 1353 GADVAAESRAKTDVARMLEQA 1373

RESULT 12

US-10-467-657-6322
; Sequence 6322, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6322
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae

US-10-467-657-6322

Query Match 2.5%; Score 103.5; DB 6; Length 2004;
Best Local Similarity 19.6%; Pred. No. 7.8;
Matches 145; Conservative 108; Mismatches 279; Indels 209; Gaps 36;
QY 144 RSREYNE-----GLDDDEEAIVTDPMDKVPKGALCVLOEG--CETEIVLVR 189
Db 712 KALREHAESIKKRLAESIGGLAERVDVAASVETAPDKAQMLLSQRVGDFDRTGKITLV 771
QY 190 GEALLKERLPKSPKQDVVVLWALAFLOVSRDAMALDDPDPFITGYEFVEEALK----- 241
Db 772 AENLTPEP-----AVMAAHELGHGFAAD-----GPAKYREELERADGNGLI 814
QY 242 -----LLQE--EGASSLAPDLR-AQIDETLEETPRVYVLELLGLPLGDDVAAKRLNGLSG 293
Db 815 RRIADAVQEGREGTGDAASVRPAAVEAVAL-----YAAQRTGGWAG 858
QY 294 VRNLTW-SVGGG-----GASALVGGITR---EKFMNEAFLRMTAAEQVDLFPVATPSNIP 343
Db 859 IENRYGVKNGLRGKIAGVLARIGALLRRVLQRLAGKAGGAMSDADVAFMLADLHGNGVE 918
QY 344 AESFEVY--EVALALVAQAFTGKPKPHLLQDADKQFOOLQOAKVMAMEIP--AMLYDTRNN 399
Db 919 GARDAPMGNGHRAVMFAAEADGAERSKESLEKRLRAETIRISGREVPPEGNLRREYKRN 978
QY 400 WEIDFGLERGLCALLIGKVDECRMWLGL-----DSEDSQY-RNPAIVEFVLENS-- 447
Db 979 -ALEYG--KSLRGPVYNKDTGREISLGRSGITELRHLDYKDAEHLQSIATAIPQIIENAVY 1035
QY 448 -----NRD--DNDDLPGLCKLLETWLAGVVF-----PRFRDTKDKKFK 483
Db 1036 IDTLPNEDLAKNGDIQG---YEVYVSGLVNGGADYTVRAAVAVSRNGNRYYYDHKLTKIE 1091
QY 484 LGDYDDPMVLSYLERVEV---QGSPL-----AAAAWARIAGAEHVK 523
Db 1092 KGN-----LLSLDRVSTTGASESKSPSGIDDKRLQLQDKDAGKGGIADPDTBAVR 1145
QY 524 ----ASAMQALQVFPSSRYTDRNSAEPKQVETVFSVDPVGNVNGRGEPCGVF--TAEAV 577
Db 1146 FSRAANIEAIGRITGKKSLDRNALKDRWDASKGIQLQFLGRRQIEDIYGGVLDGLKEYG 1205
QY 578 RPSENFTNDYATRAGVSESSVDTEVMSVADMLKEASVKILAAGVAIGLISLFSQKYF 637
Db 1206 RLSELFAGD-----ANKAVTEADKVVREWG---RLKEEDAKALA-----DLMDHATL 1249
QY 638 LK-SSSSFFQKD-----MVSSMESDVATIGS--VRADDSEALPRMDAR 677
Db 1250 AKVDADPLMRKDAQKRLDGI RTALDIADGKIEKAEAAVASAGARIAADAAYNKAQRAAD 1309
QY 678 TAENIVSKWQIKSLAFGPDHRIEMLPVLDGRMLKI-WTDRAAEATAQLGLVYDYLTKL 736
Db 1310 KAAYALEKAQE-----KHGREILADEADMRLRLFYADSEAKRA-----LRA 1352
QY 737 SVDSVTVS---ADGTRALVEA 754
Db 1353 GADVAAESRAKTDVARMLEQA 1373

RESULT 13

US-11-074-176-98
; Sequence 98, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176

```
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-98

Query Match      2.5%; Score 103; DB 7; Length 565;
Best Local Similarity 22.2%; Pred. No. 1.2; Indels 158; Gaps 30;
Matches 112; Conservative 67; Mismatches 167; Indels 158; Gaps 30;

QY 200 KSGFQDVVLVMAAFIDVSRDAMALDPPDF--ITGYEFVEEALKLLQEBGASSLAPDLRA 257
DB 123 KSKYQ-----MPLALYQIQTKERDENRPRFGLLRGEFV-----MLDGYSPAA--TRE 168

QY 258 QIDETLEETPRYVLELLGLPLGDDYAAKRLNLSG--VRNILMSVGGGG-----AS 307
DB 169 OLDQOPDDQKSAY-----KRIFKRAGVTVPVIAADSGTMGGKKNSTEFQAP 213

QY 308 ALVGG--LTRKFNWAEFLRMTAAEOVDLF-----VATPS-NIPAESEFEVYE 351
DB 214 AAIAGEDTIAATNEKGTAAANLEM--AKSIDTFKQEPPEAKELTKVATPACDTTKKJAEFLD 271

QY 352 VALALVAQAFI---GKKPHLLQADAKQFOLOQAQVM-AMEIPAMLYDTNNWEIDFGL 406
DB 272 VPATRIKVSILYIADQKVLVLIRGDKQINEVKLGHVLDADDIHEA--NTEDLKEIT-GS 328

QY 407 ERLGLCALILGKV-----DEC-----RMWLGLDSEDSQVRNP-----AIVFV-L 444
DB 329 EXG-----GVGPVNADWADKIADETVKGLYVNVVVGAGETDYQPKANLORDPKVDEFADI 384

QY 445 ENSNRDNDLGL-CKLETWLAGVVFPRFRDTKDKFKLGDYDDPMVLSYLER-----499
DB 385 RTANEGEPDVPDHLPLKFTTSIEVGH-----FKLGTYTYTKMGADFLDQNGKA 433

QY 500 VEVQGSPLAAATMARICAEHVKASAMQALQKVPFRVYTRDNRASRPKDVQETVFSVDPV 559
DB 434 QPVIMGS-----YGVGVTRM-----LSAVVEQHILTRGVAMPKEI--APFEIHIV 476

QY 560 GNNVGRDGPFGVFIABAVRP--SENPE--NDYAIRAGVS-----595
DB 477 QMKMKNEDQ--TELAELKEKKSEKVDVLYDDRKEKAGVKFADADLVGAPVRITIGCKAA 534

QY 596 -----ESSVDETTVMVSADMLK 613
DB 535 DGIVEVKRPTDEKAVMSIDELDK 558

RESULT 14
US-10-858-730-27
; Sequence 27, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgev, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858, 730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
```

```
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-858-730-27

Query Match      2.5%; Score 101; DB 6; Length 403;
Best Local Similarity 21.3%; Pred. No. 0.99; Indels 120; Gaps 16;
Matches 76; Conservative 51; Mismatches 109; Indels 120; Gaps 16;

QY 56 SSSSPATATTTATLVSLPPSIDRPER-----HVPIDIFYQVLGAQTHFLTDGIRRA 108
DB 62 SAKAEAFADITNTVRAILLEIGYDSSDKGPDGATCGVNIGI-----GAQSPDIAQGVDTA 115

QY 109 FEARV-----SKPQFGSDDALISRRQILOACETLSNPRSRREYNEGLLD 155
DB 116 HEARVEGAADPLDSQAGADQGLMFGYA---INATPELMPLPIALAHRLSRR-----163

QY 156 DEBATVITDVPWDKVPKALCVLQEGGETEIVLRVGEALLKERLPKSKQDVVLVMAFL 215
DB 164 -----LTEV---RKNGVLPLYLRPDGKTQVTI-----AYEDNVVRLDTVVISTQHAADI 209

QY 216 DVSRDAMALDPPDFITGYEFVEEALKLLQEBGASSLAPDLRAQIDETLEETPRYVLE-- 273
DB 210 DLEK--TLDPD-----IREKLVNTVLDDL---AHETLDASTVRLVNPT 248

QY 274 ---LLGLPLGDDYAAKRLNLSGVRNIL-----WSVGGGGA-----306
DB 249 GKPVLGCPMGD-----AGLTGRKIIVDITYGWARHGGGAFSGKDPKSVDRSAAYAMRW 301

QY 307 ---SALVGLITREKFNWAEFLRMTAAEOVDLFVAT-----PSNIPAESEFEVYE 352
DB 302 VAKNVVAAGLAERVEQVAY-AIGKAAPVGLFVETGTETEDPVPKIEAIGEVFDL 356

RESULT 15
US-10-467-657-2966
; Sequence 2966, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2966
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2966

Query Match      2.5%; Score 101; DB 6; Length 711;
Best Local Similarity 19.2%; Pred. No. 2.4; Indels 244; Gaps 35;
Matches 138; Conservative 98; Mismatches 237; Indels 244; Gaps 35;

QY 169 KVPKALCVLQEGGETE---IVLRVGEALLKERLPKSKQDVVLVMAFLVDSRDAMALD 225
DB 77 KIPGGF-FKREGKQSEKEILTSRIDRPIRLPLPEGYHDIQIVAMVSVDPDEIDS---132
QY 226 PPDFITGYEFVEEALKLLQEBGASSLAPDLRAQIDETLEETPRYVLELLGLPLGDDYAA 285
```

```
Db      133  IPAML-----|||:-----ALVLSVFPAGPIGA 156
Qy      286  KRLNGLSGVRNILWSVGGGASALVGLTREKFMNEAFLRWTAAB-----QVDLFVATPSN 341
Db      157  ARVGYINGV-----YVLNPTKAEAKSOLDLVVAGTS- 188
Qy      342  IPAESFEVVEVALVAQAFTGKKPHLLQDADKQFOQLQA-----KVMAMEIPAMLYDTR 397
Db      189  -----KAVLMVESEAKILPEDVMVGAVVYGHQDMQVAINAINEFADEVNPEVMDWK 239
Qy      398  ---NNWEIDFGLERGLCALLIGKVECRMWLGLDSEDSQYRNPAI VEFVLENSNRDND 454
Db      240  APETNEEL-VAKVRCIAGETIKEAFKIQKQARSAKLDEAWNAVKEALITTEETDLAANE 298
Qy      455  LPGCLKLETLWLAGV-----FPRF--RDTKDKFKLGDYDDPMVL--SYLERVEVV 503
Db      299  IKGIFKELE---ADVRSQILDGQPRIDGRDTRTVR-----PLNIQTGVLP--T 343
Qy      504  QGSPAAAAATWARI GAEHV KASAMQALQKVPFSPRYTDRNSAEPKDVQETVFSVDPVG-NN 562
Db      344  HGSALFTRGETQALAVATLGTSRDEQIIDALSGEYTDRLFMLH-----YNFPPTSTGE 395
Qy      563  VGRDGEF-----GVFTAEAVRPSNPETNDYAIRAGVSESSVDDETTVEMSVAD 610
Db      396  VGRWGA PRREIGHRLAKRALLAVLPK-EDF---SYTRKV-VSE--ITESNGSSSMAS 448
Qy      611  M-----LKEASVKILA--AGVAIGLISLFSQKYFLKSSSSFORKDMVSSMESDVATIGS 662
Db      449  VCGCLSLLSAGVPLKAHVAGIANGLI-----LEGNKFAVLT 485
Qy      663  VRADDSEALPRMD---ARTAEIVSKWKQIK-----SLAFGPDHRIEML----- 703
Db      486  DILGDEHDLGDMDFKVAGTTGEGVTALQMDIKIQGITKEIMQIALAQAKEARLHILQDKA 545
Qy      704  ----PEVLGDMKLIWTDRAE-----TAQLGLVYDYTLKLSVDSVT 742
Db      546  AVAGPOELSAAHAPRLFTMKISQDKIRDVIGKGGETIRNSITAETG-----TEIN 593
Qy      743  VSADGTRALVEATLEESACLSDLVHPENNATDVRTYTRYEVFWSKSGWKITEGSVL 799
Db      594  IAE DGT-ITTAANTQEAG-----DAKKRIEETAEVEV-----GKVEGTUV 635
```

Search completed: December 4, 2005, 12:42:48
Job time : 15 secs

(FILE 'HOME' ENTERED AT 14:25:11 ON 04 APR 2007)

FILE 'BIOSIS, CAPLUS, CABA, AGRICOLA' ENTERED AT 14:25:30 ON 04 APR 2007

L1	47 S FTN2 OR ARC6
L2	22 DUPLICATE REMOVE L1 (25 DUPLICATES REMOVED)
L3	375 S PLASTID DIVISION
L4	39 S L3 AND REVIEW
L5	31 S L4 AND ENGLISH/LA
L6	28 DUPLICATE REMOVE L5 (3 DUPLICATES REMOVED)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:20:59 ; Search time 47 Seconds
(without alignments)
1409.005 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVIGLGFQCLRLP.....YEVFWSKGWKITEGVLAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCUTUS_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	3.6	6095	2	US-09-144-085-2
2	143	3.5	2089	1	US-08-418-893D-23
3	143	3.5	2089	1	US-08-418-893D-24
4	139.5	3.4	954	2	US-09-057-969-2
5	131.5	3.2	1194	1	US-08-680-326-35
6	128.5	3.2	876	1	US-08-436-664-20
7	128.5	3.2	876	1	US-08-436-664-32
8	128.5	3.2	876	1	US-08-436-664-34
9	128.5	3.2	876	2	US-09-135-642-20
10	128.5	3.2	876	2	US-09-135-642-32
11	128.5	3.2	876	2	US-09-135-642-34
12	128.5	3.2	876	2	US-08-394-232A-20
13	128.5	3.2	876	2	US-08-394-232A-32
14	128.5	3.2	876	2	US-08-394-232A-34
15	128.5	3.2	876	4	PCT-US95-04080-20
16	128.5	3.2	876	4	PCT-US95-04080-32
17	128.5	3.2	876	4	PCT-US95-04080-34
18	126.5	3.1	2482	2	US-09-252-991A-16967
19	125.5	3.1	5087	2	US-09-144-085-1
20	122.5	3.0	2756	1	US-08-375-709-11
21	122.5	3.0	2756	1	US-08-752-929-11
22	122.5	3.0	2756	2	US-09-090-793-7
23	122.5	3.0	2756	2	US-09-231-899-7
24	121.5	3.0	680	2	US-09-252-991A-26639
25	121.5	3.0	947	2	US-09-540-236-1991
26	121.5	3.0	1220	2	US-09-540-236-3011
27	120	3.0	852	2	US-09-585-858-19

28	120	3.0	852	2	US-10-270-878-19	Sequence 19, Appl
29	120	3.0	1262	2	US-09-198-452A-97	Sequence 97, Appl
30	120	3.0	1266	2	US-09-438-185A-83	Sequence 83, Appl
31	119.5	2.9	831	1	US-08-073-384C-5	Sequence 5, Appl
32	119.5	2.9	831	1	US-08-254-359A-5	Sequence 5, Appl
33	119.5	2.9	831	1	US-08-483-043-5	Sequence 5, Appl
34	119.5	2.9	831	1	US-08-481-238-5	Sequence 5, Appl
35	119.5	2.9	831	1	US-08-471-066B-5	Sequence 5, Appl
36	119.5	2.9	831	1	US-08-484-956-5	Sequence 5, Appl
37	119.5	2.9	831	1	US-08-757-653-5	Sequence 5, Appl
38	119.5	2.9	831	1	US-08-599-491-5	Sequence 5, Appl
39	119.5	2.9	831	1	US-08-756-386-5	Sequence 5, Appl
40	119.5	2.9	831	1	US-08-823-516-5	Sequence 5, Appl
41	119.5	2.9	831	2	US-08-682-853A-5	Sequence 5, Appl
42	119.5	2.9	831	2	US-08-759-038-5	Sequence 5, Appl
43	119.5	2.9	831	2	US-08-758-314-5	Sequence 5, Appl
44	119.5	2.9	831	2	US-09-350-309-5	Sequence 5, Appl
45	119.5	2.9	831	2	US-08-520-946-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6095
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-144-085-2

Query Match	3.6%	Score 145.5	DB 2	Length 6095
Best Local Similarity	19.3%	Pred. No. 0.0036		
Matches 167	Conservative 128	Mismatches 349	Indels 221	Gaps 38
QY	70	LVSLPSPSIDRPERH-VPIPIDFYQVIGAQTHFLTDGIRAREARVSKPQFCFSDALIS	128	
Db	1642	LALLDALSPEASLVFVHLDLAQ-----LQGLSESGELPALF-----RALL	1684	
QY	129	RROILOAQCETLSNPRSRREYNEGLDDEATVTVDVFWDKVPGALCVLQEGGETEIVLR	188	
Db	1685	RPSLRKASSATRRDASALRRLSALPEARLNALVELVRGEV-AAVAGLQR-----	1734	
QY	189	VGBALLKERLPKQKQDVLMALAFDVSRDAMALDPDFITGYEFVEEALKLQEEGA	248	
Db	1735	-GEAADAQVLKELGLDLSMAVALRNLTSRTETSL-PATLVFDYPTPRAIAELLKQAF	1792	
QY	249	SSL-APDLRAQIDETLEEITPRYVLELL-----GLPLGDDVAAKRLNGLSGVRNI--LWS	300	
Db	1793	SGLQVKEARVRRRRACKDEPIATVSMACRLPGGVATPDDYWRLLAEGKDAIEGLPARWD	1852	
QY	301	-----VGGGASALVGLTREFKFNEA-----FLRWTAEEQVD	333	
Db	1853	GFEVYDDPDPBAAGSKSYAREGGFVRDIDLFANFFGISPREAQSMDDPQHRVLVLETAWEAL	1912	
QY	334	LFWATPSNIPAESFEVVEVALA-----LVAQAFIGKKPHLLQDADKQFOQLQ	380	

Db 1913 RAGVPSALSGSATGYLGSMGSDYCALHTVDLKELDGYRGSAASILSG-----R 1964
QY 381 QAKVMAMEIPAMLYTRNNWEIDFGLERGLCALLIGKVDECRMWLGDSSEDSQYRNPAT- 439
Db 1965 VAYALGLOQPAMTVDTACSSSL-VSLHLACTALROGECB-----LALAGGVTVMTSPALF 2018
QY 440 VEFV-LENSNRDDDLPGCLKLLET-----WLAG-----VFPFRFDTKDKFKLGDYVD 489
Db 2019 VEFSLKGMRSR-----GRCKSFVQADGAGWAEGCGMLLLKRLSDAQ-----RD 2063
QY 490 DPMVLSYL-----GRCKSFVQADGAGWAEGCGMLLLKRLSDAQ-----RD 2063
QY 490 DPMVLSYL-----GRCKSFVQADGAGWAEGCGMLLLKRLSDAQ-----RD 2063
Db 2064 GDRVLGVIRGSNAVDGRSQGLTAPNGPAQRVIRQALSSCGLSPEDDIDAVEAHGTGTSL 2123
QY 505 GSPLAAATMARIGAE-----HYKASAMQA-----LOKVPPSPRYTDRNSAE 545
Db 2124 GDPTEAGALAEVFGPERSPERPLYLGSSKSNLGHQAAGAGVAGVIMKWL-----SMQ 2175
QY 546 PKDQVETVPSVDVPGNNVGRDGEPPVIAEAVRPSNFETNDYAIRAGVSSSVDETTVE 605
Db 2176 HEVLKTLHAEP-SPHIGWEGS-GLSLUQEARP-----WRNRGRVRRAGVSSFGISGTNAH 2230
QY 606 MSVADMLKEA-----SVKILAAAGVAIGLISLFSQKYPFLKSSSSSFKQDMVSSME---SDVAT 659
Db 2231 IILEAPAEARREPVEAEAPALLPLVLSGRDEAAVNAQAQWAKWLEEHGEVGSVDVVR 2290
QY 660 IGSVRADSEALPRMDARTAEINIVSKWQIKSLAFG-PDHRTEMPLPEVLDGRMLKIWTD 718
Db 2291 TAAHRTHEFESASVLAASAAGAV-----EGLRALSSORPDAAVVGTAKEGGLAVLFTQ 2347
QY 719 AAEATQAG-LVYDTLLKLSVDSVTVSAD-----GTRALVEATL---EESACL---SDLVH 767
Db 2348 GQRLGMKRLYEVYVPVFAAFDEVECEALDAYLDRLGLEVVFAAAGSEBEGALLERTYTO 2407
QY 768 PENNATDVRTYTRTEVFWFSKSWK 792
Db 2408 PGLFALEVALYRQ-----WESWGLK 2427

RESULT 2
US-08-418-893D-23
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-418-893D-23

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred. No. 0.00096;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

QY 73 LPPSIDRPERHVPIDFYQVLGAQTHFLTDGIRRAPEARVSKPP-----QFGSDAL 126
Db 197 LPNALDK-----LGKFGTGTGPVMSVLGDKIAANILAQTAKVPSIPWSGSGFGPDGP 250
QY 127 ISRRQLQAACTLSNPSRRREYNEGLD--DEEATVITDPWDKVPALCVLEGG--- 181
Db 251 LQ-----ADLTEGTIPMEIFNKLVTSADEAVIVANKIGHEN--GIMIKASEGGGK 301
QY 182 -----ETEIVLR-----VGEALLKERLPKQDVVLVMAFLDVSRDAMALDP 226
Db 302 GTRFVNEADLRNAPVQVNEVIGSIFLMQLCKNARHIEVQIVG-----DOHGNVALNG 357
QY 227 PDFITGYEFVEEALKLQEBEGASSLAPDLRAQIDETLEETIPRYVLELLGLPLGDDYAAK 286
Db 358 RDCSTQRRF-----QKIFEEGPPSIVP-----KETFHEM-----ELAAQ 391
QY 287 RLNLGSGVRNLSWVGSGGASALVGLITREKFNNEAPFLMTAAEQVDLFVA---TPSNIP 343
Db 392 RL-----TONIGYQ--GAGTVEYLNAADNKKF---FLELNRLQVEHPVTEGITGANLP 441
QY 344 AES-----FEVYEVALALVAQAFIGKPKHLLQDADKQFOOLOQAQKNAMEIPA--- 391
Db 442 ATOLQVAMGIPUNIPDIRRLYGREDAYGTDG---IDFLQERYREL--DSHVIAARITANP 498
QY 392 -----MLYDTRNNW--EIDFGLERGLCALLIGKVDECRMWLGDSSEDSQY--- 434
Db 499 DEGFKPTSGSIERIKFQSTPNVWGYFVGANGI-----HEFADSQEGH 542
QY 435 ---RNP-----AIVEFVLENSNRDD--NDDLPGCLKLLET-----WLAGVVP 472
Db 543 LPAKGPNRQAKALVLALKEMEVRGDIRNSVEYLKLETEAFKNTIDTSLDGLII-- 600
QY 473 RFRDTHDKKFKLGDYDDPMVLSYLERVEVQGSPLAAATMARIGAHHKASAMQA--- 529
Db 601 -----KEKSVKV-----EMPSHL-----VVVGAAPFAFEHVKVATEEVKESPRKQVS 644
QY 530 -----LOKVPPSPRYTDRNSAEPKQVQET----- 552
Db 645 TAGIPGINSFNIEVAYLDTKYPFHVERISPDVTRPFLDGNITDVEVTQTAEGALLATFG 704
QY 553 -----VFSVD--PVGNNVGRDGE-----PGVFIAEAVRPS-----ENFTNDYAIRAGVSES 597
Db 705 ETRHIFGMBEPLGLRLSLDGATVLMPTIFDPSELRTDVTGKVVRYLQDNGATVEAG--- 760
QY 598 SVDETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYPFLKSSSSSFKQDMVSSME--SD 656
Db 761 ---QPYVEVEAMKMI--MPIKATESG-----KITHNLASGVSISAGDGLASLELD 806
QY 657 VATIGSV-----RADDSEALPRMDARTA--ENIVSKWQIKSLAFGPDHRIEMLEVLDR 710
Db 807 PSRVKKIETFSKGLDIMEKVDLEPQKAVNVLIS-----GFNLD-----PEAVAQ 852
QY 711 MLKIWTDRAAETAQLGLVYD-YTLKLSVDSVTVSADGTRALVEATLEESACLSDLVHPE 769

Db 853 AIDSATDSSAAADLLVQLDEFYRVESQFDGV-IADDVVTITKANTE-----TLDVWSE 907
QY 770 NNA 772
Db 908 NLA 910

RESULT 3

US-08-418-893D-24
; Sequence 24, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-418-893D-24

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred.No. 0.00096;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;
QY 73 LPSPIDRPRHRVPIPIIDFYQVLGAQTHFTDGTIRRAFEARVSKPP-----QFGPSDDAL 126
Db 197 LPNALDK-----LGKFTGPTGPMVSVLGDKIAANILAOATKVPISIPWSGSGFGGPDGP 250
QY 127 ISRRQLQACSTLGNPRSRREYNGLD--DEEATVITDVPMDKVPGLCVLQEGG---181
Db 251 LQ-----ADLTEGTIPWEIFNKGVLTSDEAVIVANKIGWEN--GIMIRASEGGGK 301
QY 182 -----ETEVLV-----VGEALLKERLPKPKQDVVLVMAFLDVRDAMALDP 226
Db 302 GIRFVDNEADLRNAPVQVSNEVIGSPFIPLMQLCKNARHIEVQIVG-----DOHGNVALNG 357

QY 227 PDFITGVFEVEALKLQEGCASSLAPDLRAQIDETLEETIPRYVELLGLPLGDDYAAK 286
Db 358 RDCSTORRF-----QKIPBEGPPSIVP-----KETTHEM-----ELAAQ 391
QY 287 RLNGLSGVRNILWSVGGGASALVGLTREKPMNEAPLRTAAEQVDFVA---TPSNIP 343
Db 392 RL-----TONIGYQ--GAGTVEYLYNAADNKF--FLELNPRLQVHPVTEGITGANLP 441
QY 344 AES-----FEVYEVALVAQAFIGKKPHILLQADAKQFQQQAKVMAEIPA--- 391
Db 442 ATQLQVAMGIPLFNIPDIRRLYGREDAVGTDP--IDFLQERYREL--DSHVIAARITAENP 498
QY 392 -----MLYDTRNNW--EIDFGLBERGLCALLIGKVDCEMRMLGLDSEDSQY--- 434
Db 499 DEGFKPTSGSTIERIKFQSTPNVMGYFVSANGGI-----HEFADSQFQGH 542
QY 435 ---RNP-----AIVFVLENSNRDD--NDLPLGLCKLLET-----WLAGVVFPP 472
Db 543 LFAKGNREQARKALVIALKEMEVRGDIRNSVEYLVKLETEAFKKNITDTSWLDGII-- 600
QY 473 RFRDTKDKFKLGDYDDPMVLSYLERVEVVGQSPALAAATMARIGAEHVKASAMQA--- 529
Db 601 -----KEKSVKV-----EMPSHL-----VVGAAPVKAPEHVKVATEEVKESPRKGQVS 644
QY 530 -----LQKVFPERYTDRNSAEPKDVQET----- 552
Db 645 TAGIPGINSFNIEVAYLDTKYPFHEVERISPDVYRFTLDGNTIDVEVTQAEALIAATFGG 704
QY 553 ---VFSVD--PVGNVVRDGE---PGVFIAEAVRPS-----ENFEINDAIRAGVSES 597
Db 705 ETHRIFGMDPLGLRLSLDGAATVLMPTIFDPSELRDTVTGKVRVYLDQNGATVEAG---- 760
QY 598 SVDETTVMESVADMLKEASVKILAAQVAIGLISLFSQKYLKSSSSSFQKDMVSSME--SD 656
Db 761 ---QPYVEVEAMKMI--MPIKATESG-----KITHNLSAGSVISAGDLLASLELKD 806
QY 657 VATIGSV-----RADDSALPRMDART--ENIVSKWQKIKSLAFPGPDHRIEMLPEVLDR 710
Db 807 PSRVKKTETFGKLDIMESKVDLEPQKVMNVL-----GFNLD-----PEAVAQQ 852
QY 711 MLKIWTDRAAETAQGLGYD--YTLKLSVSVTVSADGTALVTEATLEESACLSDLVHPE 769
Db 853 AIDSATDSSAAADLLVQLDEFYRVESQFDGV-IADDVVTITKANTE-----TLDVWSE 907
QY 770 NNA 772
Db 908 NLA 910

RESULT 4

US-09-057-969-2
; Sequence 2, Application US/09057969
; Patent No. 6013451
; GENERAL INFORMATION:
; APPLICANT: WONG, VICTOR THI WONG
; APPLICANT: PHANG, SENG MENG
; APPLICANT: TAN, TIEN CHVE
; TITLE OF INVENTION: BACILLUS STEAROTHERMAPHILUS DNA
; TITLE OF INVENTION: POLYMERASE I (KLENOW) CLONES INCLUDING THOSE REDUCED TO
; TITLE OF INVENTION: 3'-To-5' EXONUCLEASE ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,969
; FILING DATE: 09-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 674507-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; US-09-057-969-2

Query Match 3.4%; Score 139.5; DB 2; Length 954;
Best Local Similarity 22.0%; Pred. No. 0.00054;
Matches 175; Conservative 132; Mismatches 315; Indels 173; Gaps 44;

QY 99 HELTDCIRRAFEARVSKPPQFGSDALISRRQILQAACETLSNPRSREY-NEGILLDE 157
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 HYEADDIIIGTMAAREE---GFVAVKISGDRDLQ-----LASPQVTVEITKKGITDIE 234
QY 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
235 SYTPEVVEKYGLTPEIQIVDLKGLMGKSDNIPGVGIGKGTAVKLLQFGTVENYLSAI 294
QY 191 EALLKERLPKSKFO--DVLVWMAFLDVSRA---MALDPPDPIITGYEFVEALKLQOE 245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 DEIKGEKLENRQVRDLAL--LSKQLAAICRQAPVELTLD--DIVYKGEDREKVVAFQOE 351
QY 246 EGASSIAPDLRAQIDE-----TLEEITPRYVLELLGL-----PLGDDYAAREKLN- 290
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
352 LGQFSLDKQAVQTDGEKPLAGMDFAIADSVTDEMLADKAALVVEVGDVGHYHPIVGI 411
QY 291 -LSGVNIIW-----SVGGGGSALVGLTRKFM-----NEAFIRMFAEQV-D 333
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
412 ALANERGRFLLPETAVADPKFLWLGDETCKTKTFDMSKRAAVALKGKIELAGVGVVFD 471
QY 334 LFVATPSNIPAS-----FEVVEVALALVAQAFIGK-KPHLLQADKQFOOL--QOAK 383
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
472 LLLAAYLLDPAAGDVAAVAKMHOYEAVRSDEAVYKGAARTVPDEPTLAQLVRKAAA 531
QY 384 VMAMEIPAMLYTRNNWE-IDFGLERGLCALLIG-----KVDECRM-WLGLD-SEDSQY 434
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
532 IWALEPLMDLRRNEQDRLTLEHALAGILANMEFTGVKVDTKLEQMGAEITLQLO- 590
QY 435 RNPAIVEFVLENSRDNDDLPGLCKLLETWLAGVVFPRPRTDKKKFKLGDYDDPMVL 494
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
591 ---AVERRIYELAGQEFNINSP---KQLGT---VLFDKQLPLVKTKTKTG-YSTSADVL 639
QY 495 SYL-ERVEVVOQSPLAAAATMARIAGEHVKASMQALQKFFPSRYTRDSAEKQVET- 552
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
640 EKLAPHHEIVE--HILHYRQLGKLOSTYTB-GLLKVVHPVTGKVHTMFNQA-----LTQTG 692
QY 553 -VFSVDPVGNVGRDGPVGTAEAVRPSH-----NFETNDYAIRAGVSSSVDETT 603
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
693 RLSSVEPNLQNIPIRLEGRKTRQAVFPSEPDWLIIPAADYSQIELRVLAHAE---DDNL 749
QY 604 VE-----MSVADMLKE---ASVKILAAAGVAIGLISLFQSKYFLKSSSSFORK 647
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 IEAFRRLWDIHTKTAMDIPHVSEEDVTANMRQAKAVNFGIVYGISD-YGLAQLNITRK 808
QY 648 DMVSSMESDVATIGSVRADDSALFRMDARTAEATIVSKWKQIKSLAFGPDHRIEMLPVL 707
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
809 EAAEFIERVAFSPGVK-----QYMDNIVQOE-AKQGYVTTLLHRRRYLPDIT 855
;

QY 708 DGRM-LKIWTDRAA-ETAQLGLVYDYTLKLSVD-SVTVSADGTRA-----LVEA 754
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
856 SRNFVTRTFAERTAMTPIQGSAA--IIKAMIDLVSREERLQARLLQLQGHDELILEA 914
QY 755 TLESACLSDLVHPE 769
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
915 PKEIIGRLCLRV-PE 928

RESULT 5
US-08-680-326-35
; Sequence 35, Application US/08680326
; Patent No. 5925733
; GENERAL INFORMATION:
; APPLICANT: ROSE, TIMOTHY M.
; APPLICANT: BOSCH, MARNIX
; APPLICANT: STRAND, KURT
; APPLICANT: TODARO, GEORGE J.
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; NUMBER OF SEQUENCES: 152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,326
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-680-326-35

Query Match 3.2%; Score 131.5; DB 1; Length 1194;
Best Local Similarity 19.0%; Pred. No. 0.0049;
Matches 168; Conservative 122; Mismatches 297; Indels 297; Gaps 44;

QY 48 SDNFNSTDSS-----SSPATATTTATLVSLPPSTDTRERHV--PIPIDFYQVLGAQ---TH 99
Db 205 NDATLNGDKNAFTGTSFKASPSFRFEV---IERTDVTYYDTQPCAFYRVYSPSSKFTN 261
QY 100 FLTDGIR--RAFEARVSKPPQFGSDALIS-----RRQI---- 132
Db 262 YLCDNFHPELKTYEGRVDATTRFMDNPGVSGVQLKPGVDGERVVRPASQLTUSD 321
QY 133 LQAAC-----ETLSNPRSREYNEGILLDE-----EATVITDVPWDKVPGLALC 175
Db 322 VEIDCWSDNLQAIAPNDSDWPDYKLLCFDIECKSGSNELAFPDATHLEDL---VIQISC 377
QY 176 VL-----QEGGETEIVLRVGEALLKERLPKSKFO-----DVLVWMAFLDVSRD 220
Db 378 LLYSIPRQSLHILLFSLGSCDLFPQYVQEMKDAGLPEPTVLFPDSEFELLIAFMTLVKQ 437
```


Db 662 LAHIAE---DDNLIEAFRRGLDIHTKTAMDIFHVSEEDVTANRRQAKVNFIVGISD 718
QY 635 KYFLKSSSSFFQRKQMVSSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQNLTITKEAAEFIERFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLEPEVLDRGM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMNTPIQGSAAD-IKKAMIDL SVLRERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 7

US-08-436-664-32
; Sequence 32, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,664
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/394,232
; FILING DATE:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A.
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:

INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-436-664-32

Query Match 3.2%; Score 128.5; DB 1; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;

Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
QY 99 HFUTDGIIRAFARVAKPPQFGSDDALISRRQILQACETLNSRREY-NEGLLDDE 157
Db 107 HVEADDIIIGTMAARE---GFAVKVISGDRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PMDKVPG-----ALCVLQGGTETIVLRVG 190
Db 159 SYTPETVWEKYGTPEQIVDLKGLMGDKSDNIPGPGIGBKTAVKLLQKQFQTVENVLASI 218
QY 191 EALLKRLPKSPKQ--DVLVMAALAFIDVSRDA---MALDPPDPFITGYEFVEEALKLQE 245
Db 219 DEIKGKLENLRQYRDLAL-LSKQLAAICRDPVELTLD--DIVYGEDREKVVALLQE 275
QY 246 EGASSIAPDLRAQIDE-----TLEEITPRYVLELGL---PLGDDYAAAKRLNG- 290
Db 276 LGFQSFLLDKMAVQTDGEKPLAGMDFADISVTDMLADKAALVVEVGDNYHHAIPVGI 335
QY 291 -----LSGVRNILWSVGGGASALVUGLITREKPMNEAFLMTAAE----- 330
Db 336 ALANERGRFRLPETALADPKFLAW-----LGDETKKTKTFDSKRAAVALKWKGI 385
QY 331 ----QVDLFVATPSNIPAES-----PEVYEVALALVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVVDLLLAAYLLDPAAGDVAACVAKHQVEAVRSDEAVYGGAKRTVPDEPTLAE 445
QY 371 DADKQFQQLQQAQKAMAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HILAR-----KAAAIWALEEPLMBELRNEQDRLLTELEQPLAGILANWFTGVKVDTKRL 500
QY 424 -WLGLD-SEDSQYRNPAIVFVLNENRDNDDLLPGLCKLLETWLAGVWPPRFRDTKDKK 481
Db 501 EQMGAELEQLQ-----AVERRIYELAGQEFNINSP---KQLGT---VLEFDKLQLPVLKK 549
QY 482 FKLGDDYDDPMVLSYL-ERVEVVGQSPAAAAATWARGAHSVKAASAMQALQKVPSPRYTD 540
Db 550 TKTG-YSTSADVLEKLAHPHEIVE--HILHYRLQGLQSTYIE-GLLKVHPVTKVHTM 605
QY 541 RNSAEPRKDVQET--VFSVDPVGNVGRDGPVFIABAVRPS-----NFTNDYAI 590
Db 606 FNOA-----LTQTGLSSVEPNLQNIPIRLEGRKIRQAFVPSFEDWLI FAADYSQIELRV 661
QY 591 RAGVSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLPSQ 634
Db 662 LAHIAE---DDNLIEAFRRGLDIHTKTAMDIFHVSEEDVTANRRQAKVNFIVGISD 718
QY 635 KYFLKSSSSFFQRKQMVSSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQNLTITKEAAEFIERFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLEPEVLDRGM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMNTPIQGSAAD-IKKAMIDL SVLRERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 8

US-08-436-664-34
; Sequence 34, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive

CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,664
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:

INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-436-664-34

Query Match 3.2%; Score 128.5; DB 1; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDCIRRAFEARVSKPQFGSDALISRRQILOAACETLSNPRSRREY-NEGLLDDE 157
DB 107 HYEADDIIGTMAARE--GFAVKVISGDRDLTQ-----LASQVTVETIKKGITDIE 158
QY 158 EATVITDV-----PMDKVFG-----ALCVLQEGGETEIVLRVG 190
DB 159 SVTPETVVEKYGLTPEQIVDLKGLMGKSDNIPGVPGIGKETA VKLLKQGTVENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVWALAFDVSRA---MALDPPDFTTGYFVEALKLQ 245
DB 219 DEIKGEKLENIRQYRDAL-LSKQLAAICRDAPELVETLD--DIVYKGEDREKVALFOE 275
QY 246 ECASSIAPDLRAQIDB-----TLEETTPRVVLELGI---PLGDDVAAKELNG- 290
DB 276 LGFQSLDKMAVQTDGEKPLAGMDPAIDSVTDENLADKAALVBEVGVGNVHHAPIVGI 335
QY 291 -----LSGVNRLMSVGGGASALVGLTRKFKFNEAEFLRMTAAE---- 330
DB 336 ALANERGRFLPETALADPKFLAW-----LGDETKKTKMTFDSKRAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----FEVEVALALVAQAFIGK-----KPHLIQ 370
DB 386 ELRGVVFDDLLAAYLLDPAQAAGDVAAVAKMKQYEAVRSDEAVYKGAKRTVPDEPTLAE 445
QY 371 DADKQFQQLQAKVWAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KUDECRM 423
DB 446 HJAR-----KAAAIWALEEPLMEFLRNEQDRLLTELEQPLAGILANMEFTGVKVDTKKL 500

QY 424 -WLGLD-SEDSOYRNPALIVEFVLENSNRDDNDLPLCLCKLLETWLAGVVPFRDTPKDKK 481
DB 501 EOMGAELTEQLQ-----AVERRIYELAQEFNINSP---KQLGT-----VLPDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVOGSPLAAAAATMARIGAHEVVKASAMQALQKVPFSRYTD 540
DB 550 TKTG-YSTSADVLEKLAPHHEIVE--HILHRQLKQLQSTYIE-GLLKVVHPVTGKVHTM 605
QY 541 RNSAPKDVQET--VFSVDPVGNVGRDGPVFIAEAVRPS-----NFTNDYAI 590
DB 606 FNQA-----LTQTGRLLSSVEPNLQIPRIEGRKIRQAFVPSEPDWLIFAADYSQIELRV 661
QY 591 RAGVSESSVDETTVE-----MSVADMLKE---ASVKILAAQVAGLISLFSQ 634
DB 662 LAHTAE---DDNLTEAFRRGLDIHTKTAMDFHVSEEDVTANMRQAKAVNFGIVYGISD 718
QY 635 KYFLKSSSSSFORKDWMSSMESDVATIGSVRADDSSEALPRMDARTAEINIVSKWKIKSLAF 694
DB 719 -YGLAQNLNITRKEAAEFIERYPASFPQGVK-----QYMDNIVQE-AKQKGYVT 764
QY 695 GPDHRIEMPEVLDRM-LKIWTDRAA-ETAQLGLVYDYDTLLKLSVD-SVTVSADGTRA- 750
DB 765 TLLHRRRYLPDITSNENFVRSFAERTAMNTPIQSSAAD-IKKAMIDLVSRLREERLQAR 823
QY 751 -----LVEATLESACLSDLVHPE 769
DB 824 LLLQVHDELILEAPKBEIERLCRLV-PE 850

RESULT 9
US-09-135-642-20
Sequence 20, Application US/09135642
Patent No. 6066483
GENERAL INFORMATION:
APPLICANT: RIGGS, MICHAEL G.
APPLICANT: SIVARAM, MATHOOR
APPLICANT: TUDOR, STAKLA D.
TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
TITLE OF INVENTION: STEAROTHERMOPHILUS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 9880 Campus Point Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,642
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:

```
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-09-135-642-20

Query Match
Best Local Similarity 3.2%; Score 128.5; DB 2; Length 876;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDCIRRAFAEAVSKPPQFGSDDALISRRQILQAACETLSNPRSRREY-NEGILLDDE 157
Db 107 HYEADDILGTMAARAERE---GFAVKVISGRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PMDKVPQ-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTPETVVEKYGLTPQIVDLKGLMGDKSDNIPGVPGIGKTAVKLLKQFQTVENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVLMALAFLDVSRDA---MALDPPDFITGYEFVEEALKLQOE 245
Db 219 DEIKGEKLENLQRYRDLAL-LSKQLAAICRDAPEVELTLD--DIVYKGEDREKVVAFQOE 275
QY 246 EGASSIAPDLRAQIDE-----TLEETPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDGEXPELAGMDFAIADSVTDEMLADKAALVVEVGDNYHHPIVGI 335
QY 291 -----LSGVRNLTWVGSGGASALVGGITREKFMNEAFIRMTAAE----- 330
Db 336 ALANERGRFELRPETALADPKFLAW-----LGDETKKKTMPDRAAVALKWKGI 385
QY 331 -----QVDLPVATPSNIPAES-----FEVVEVALVAQAFIGK-----KPHILQ 370
Db 386 ELRGVVFDLLLAAYLDPAQAAGDVAVAKMHQYEAVRSDAEAYVYGAKRTVPDEPTLAE 445
QY 371 DADKQFQQLQQAQVMAWMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HILAR-----KAAAIWALEPFLMDELRRNEQDRLLITELQPLAGILANMEFTGVKVTKRL 500
QY 424 -WLGLD-SEDSQVRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRPRDKDK 481
Db 501 EQMGABELTEQLQ---AVERRIYELAGQEFNINSP--KQLGT---VLFDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVGQSPLAATAATMARIGAEHVKASAMQALQKVPSPRYTD 540
Db 550 TKTG-YSTSADVLEKLAHPHEIVE--HILHYRLQGLQSTYIE-GLLKVVHPVTVGKVHTM 605
QY 541 RNSAEPKDYQET--VPSVDPVGNVGRDGEPPGVFIAEAVRPSB-----NFTNDYAI 590
Db 606 FNAQ-----ITQGRLSVSEPNLQNIPIRLEEGKIRQAQFVPSPEPDWLIIFAADYSQELRV 661
QY 591 RAGVSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLFSQ 634
Db 662 LAHIAE---DDNLIEAFRRGLDITHTKAMDIFHVSEEDVTANRRQAQAVNFIVVIGISD 718
QY 635 KYFLKSSSSFORQDMVSSMESDVATTGSRADDSSEALPRMDARTANIVSKWQIKSLAF 694
Db 719 -YCLAQNLNITRKEAAAEFTERYPASFPQVK-----QYMDNIVQOE-AKQKGYVT 764
QY 695 GPDHRIEMLEPEVLDGRM-LKIWTDRAA-ETAQGLGVYDYTLKLKSYD-SVTVSADGTRA- 750
Db 765 TLLHRRRYLPDITSRNPNFNSPAERTAMNTPIQSSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLESACLSLVHPE 769
Db 824 LLLQVHDELILAPKEIERLCFLV-PE 850
```

RESULT 10

```
US-09-135-642-32
; Sequence 32, Application US/09135642
; Patent No. 6066483
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,642
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/394,232
; FILING DATE:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-09-135-642-32

Query Match
Best Local Similarity 3.2%; Score 128.5; DB 2; Length 876;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDCIRRAFAEAVSKPPQFGSDDALISRRQILQAACETLSNPRSRREY-NEGILLDDE 157
Db 107 HYEADDILGTMAARAERE---GFAVKVISGRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PMDKVPQ-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTPETVVEKYGLTPQIVDLKGLMGDKSDNIPGVPGIGKTAVKLLKQFQTVENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVLMALAFLDVSRDA---MALDPPDFITGYEFVEEALKLQOE 245
Db 219 DEIKGEKLENLQRYRDLAL-LSKQLAAICRDAPEVELTLD--DIVYKGEDREKVVAFQOE 275
QY 246 EGASSIAPDLRAQIDE-----TLEETPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDGEXPELAGMDFAIADSVTDEMLADKAALVVEVGDNYHHPIVGI 335
```


Db 719 -YGLAQNINIRKRAAEFIERYFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLPEVLGRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMTPIQSSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 12
US-08-394-232A-20
; Sequence 32, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, NATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-394-232A-20

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
QY 99 HFLTIGIRAFARVSKPPQFGSDALISRRILQACETLSNPRSRREY-NEGILLDDR 157
Db 107 HVEADDIIGTMAAREE---GPAVKVIGSDRLDTQ-----LASPQVTVETIKKGITDIE 158
QY 158 EATVITDV-----PWDKVPK-----ALCVLQEGGETEIVLRVG 190

Db 159 SYTPETVWKYGLTTPQIVDLKGLMGDKSNIPGVGIGEKTAVKLLKQFTVENVLASI 218
QY 191 EALLKERLPSFKQ--DVLVLMALAFLDVSRDA---MALPPDPFITGYEFVEEALKILQE 245
Db 219 DEIKGEKLENLQYRDLAL-LSQLAAICRDAPVELTLD--DIVYKGEDREKVVALLQE 275
QY 246 EGASSIAPDLRAQIDE-----TLBEITPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDGEKPLAGMDFADIAVSDTDEMLADKAALVVEVVGDNVHHAPIVGI 335
QY 291 -----LSGVNINILSVGGGASALVGLLTFEEKEMNEAFLRWTAEE----- 330
Db 336 ALANERGRFELRPETALADPKFLAW-----LGDETKKTKTMDFSKRAAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----FEVYEVALALVAQAFICK-----KPHLLQ 370
Db 386 ELRGVVFDLILLLAAYLLDPAAGDAVAARKHQVEAVRSDEAVYKGNKTRVTPDEPTLAE 445
QY 371 DADKQFOQLQOAKMAMBIAMLYDTNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HLLAR-----KAAAIWALEEPLMDLRERQDRLLLEQLAGILANMEFTGVKVDTKRL 500
QY 424 -WLGLD-SEDSOYENPAIVEFVLENSNRDDNDLPGCLKLETLWLAGVVFPRFRDTKDKK 481
Db 501 EQMGABLTELQ-----AVERRIYELAGOEFNINSP---KQLGT-----VLFDKLQLPVLKK 549
QY 482 FKLGDYVDDPMVLSYL-ERVEVVQGSPLAAAAATMARIGAETHVAKASAMQALQKVPFSRYTD 540
Db 550 TKTG-YSTSADVLEKLAPHHEIVE--HILHYRLQKLGQSTYIE-GLLKVHPVTKVHTM 605
QY 541 RNSAEPKDVQET--VFSVDPVGNVGRDGBGVFIABAVRPE-----NFETNDYAI 590
Db 606 FNQA-----LTQTGLSSVEPNLQNIPIRLEGRKIRQAFVSPEDWLIFAAIDYSQIELRV 661
QY 591 RAGVSESSVDETTE-----MSVADMLKE---ASVKILAAVAILGLISLFSQ 634
Db 662 LAHIAE---DNLIEAFRRGLDITHTKAMDIFHVSEEDVTANMRQAKAVNFVIGYISD 718
QY 635 KYFLKSSSSFORQDMVSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWQIKSLAF 694
Db 719 -YGLAQNINIRKRAAEFIERYFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLPEVLGRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMTPIQSSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 13
US-08-394-232A-32
; Sequence 32, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, NATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-394-232A-32

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDGIRAFARVSKPQFGSDDALISRRILQACETLSNPRRRY-NEGLLDDE 157
DB 107 HYEADDIIIGTMAAREE---GPAVKVIGSDRLTQ-----LASPQVTVEITKGGITDIE 158
QY 158 EATVITDV-----PMDKVPG-----ALCVLOEGGETEIVLRVG 190
DB 159 SYPTETVVEKYGLTPQIVDLKGLMGDKSDNIPGVPGIGKTAVKLUKQGTVENVLASI 218
QY 191 EALLKERLPKSFQ---DVLVLMALFLDVSRA---MALDPPDFITGYEFVEBALKILQE 245
DB 219 DEIKGEKLENLRQYRDAL-LSQLAAICROAPVELTD--DIVYKGEDEKVVAFQE 275
QY 246 EGASSLAPDLRAQIDE-----TLEETTPRYVLELLGL---PLGDDYAAKRLNG- 290
DB 276 LGFQSPDKMAVQTDGKPLAGMDPAIADSVTDEMLADKAALVVEVVGDNVHHPVIGI 335
QY 291 -----LSGVRNILWSVGGGASALVGGLTREKFMNEAFLRMTAAE----- 330
DB 336 ALANERGRFLPETALADPKFLAW-----LGDETCKTMTFDSKRAAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----PEVVEVALALVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVVFDLLLAAYLLDPAQAGDVAAVAKHQYEAVRSDAEVYGGAKRTVPDEFTLAE 445
QY 371 DADKQFQQLQQAQWAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
DB 446 HLAR-----KAAAIWALEBPLMDELRRNEQDRLLTELEQPLAGILANMEFTGVKVDTKL 500
QY 424 -WGLD-SEDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVVPFRFDYDKK 481
DB 501 EQMGAELEQLQ-----AVERIYELAGQEFNINSP---KQLGT-----VLFDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVQGSPLAAAAATMARIGAEHVKASAMQALQKVPESRYTD 540
DB 550 TKYG-YSTADVLEKLAPHHEIVE--HILHYQLQKLGQSTYIE-GLLKVVHPVTGKVHTM 605
QY 541 RNSAEPKVOQET--VFSVDVPGNVGRDGEPPGVFTIAEAVRPS-----NFTNDYAI 590
;
606 FNQA-----LTQTGRLLSVPEPNLQNIPIRLEGRKIRQAFVPSEPDWLIIFAADYSQIELRV 661
QY 591 RAGVSESSVDETTVE-----MSVADMLKE---ASVKILAAAGVAIGLISLFSQ 634
DB 662 LAHTAE---DDNLIEAPRRGLDIHTKTAMDIFHVSEEDVTANMRQAKAVNFGIVYGISD 718
QY 635 KYFLKSSSSSFORKWVSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
DB 719 -YGLAQNLNITRKEAAEFIERYPASFGVK-----QYMDNIVQE-AKOKGYVT 764
QY 695 GPDHRIEMLPVLDGRM-LKIWTDRAA-ETAQLGLVYDVYTLTKLSVD-SVTYSADGTRA- 750
DB 765 TLLHRRYLPDITSRNFVRFAERTAMNTPIQGSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
DB 824 LLLQVHDELIILEAPKEIERLCRLV-PE 850

RESULT 14
US-08-394-232A-34
; Sequence 34, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-394-232A-34

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
```

Matches	170;	Conservative	135;	Mismatches	302;	Indels	201;	Gaps	43;
Qy	99	HFUTDGIRRAFEARVSKPPOFGSDDALISRRQILOACETLSNPSRRREY-NEGLLDDE	157						
Db	107	HYEADDIIIGTMAARAERE--GFAVKVISGRDLTQ----	158						
Qy	158	EATVITDV-----PWDKVPG-----ALCVLOEGGETEVLVRG	190						
Db	159	SYTPETVVEKYGTPEQIVDLKGLMGKSDNIPGVGIGEKTAVKLLKQFGTGVENVLASI	218						
Qy	191	EALLKERLPKSFQ--DVLVLMALAFLDVSRDA--MALDPDPFITGTFYFVEEAKLILQE	245						
Db	219	DEIKGEKLENLRQYRDAL-LSQLAAICRDAPVELTLD--DIVYKGEDREKVVALLFQE	275						
Qy	246	EGASSIAPDLRAQIDE-----TLBEITPRYVLELLGL---PLGDDYAAKRLNG-	290						
Db	276	LGFQSFLDKMAVQTDEGEKPLAGDMDFATADSVTDEMLADKAALVVEVVGVDYHPIAVGI	335						
Qy	291	-----LSGVNRNLWSVGGGASALVGGLTREKFWNEAFLRMTAAE----	330						
Db	336	ALANERGRFLRPETALADPKFLAW-----LGDETCKKTMFDSKRAVALKWKGI	385						
Qy	331	-----QVDLFVATPSNIPAES-----FEYVEVALVAQAFTGK-----KPHLLQ	370						
Db	386	ELRGVPPDLLLAAYLLPQAAGDVAAVAKMHQYEAVRSDVAVYKGAKRTVPDEPTLAE	445						
Qy	371	DADKQFOOLOQAQKVMAMEIPAMLYDTNNWE-IDFGLERGICALLIG-----KVDECRM	423						
Db	446	HLAR-----KAAAIWALEEPLMDBLRNEQDRLLTELEQPLAGILLANMEFTGVKVDTKRL	500						
Qy	424	-WLGLD-SEDSQYRNPAIVFVLNSNRDDNDLPLGLCKLLETWLAGVVFPRFRDTKDK	481						
Db	501	EQMGAELEQLQ---AVERRIYELAQGFNINSP---KQLGT---VLPDKQLPVLKK	549						
Qy	482	FKLGDYDDPMVLSYL-ERVEVVGSGPLAAAATWARIAGAEHVKSAMQALQKVPSPRYTD	540						
Db	550	TKTG-YSTSDVLEKLAHPHHEIVE-HILHYRQLGKLOSTVIE-GLLKVVHPVTGKVHTM	605						
Qy	541	RNSAEPKDVQET--VFSDVPGVNVNDRDGEQVPIAEAVRPE-----NFETNDYAI	590						
Db	606	FNQA---LTQTGRLLSVEPNLQNIPIRLEBGRKIRQAFVSEPDWLIIPADYSQIEURV	661						
Qy	591	RAGVSESSVDTEVVE-----MSVADMLKE---ASVKILAAAGVAILSLFSQ	634						
Db	662	LAHIAE---DONLIEAFRRGLDIHTKTAMDIFHVSEEDVTANMRRAQKAVNFGIVYGISD	718						
Qy	635	KYFLKSSSSFORKDMWSMESDVATIGSVRADDSREALPRMDARTAEINIVSKWQIKLSLAF	694						
Db	719	-YGLAQNLNITRKEAAEFIERYPASFPGVK-----QYMDNITVQE-AKQKGYVT	764						
Qy	695	GPDHRIEMLPEVLDGRM-LKIWTDRAA-ETAQLGLVDYDTLLKLSVD-SVTVSADGTRA-	750						
Db	765	TLLHRRYLPDITSRNFNVRSFAERTAMNTPIQSSAAD-IKKAMIDLSVRLREERLAR	823						
Qy	751	-----LVEATLESSACLSLDVHPE	769						
Db	824	LLLOVHDELILEAPKEIERLCRIV-PE	850						

```

RESULT 15
PCT-US95-04080-20
; Sequence 20, Application PC/TUS9504080
; GENERAL INFORMATION:
;
; APPLICANT:
;
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
;
; TITLE OF INVENTION: STEAROTHERMOPHILUS
;
; NUMBER OF SEQUENCES: 34
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
;

```

Search completed: December 4, 2005, 12:29:59

•

Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:39:31 ; Search time 233 Seconds

(without alignments)
2425.443 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVGIGLSPFLCRUP.....YEVFWSKGWKGITGVSLAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4063	100.0	801	2	Q7XAS0 ARATH
2	4052	99.7	801	2	Q7XAR9 ARATH
3	4051	99.7	801	2	Q9FTG9 ARATH
4	1775.5	43.7	760	2	Q62729 ORYSA
5	1775.5	43.7	760	2	Q7PC78 ORYSA
6	1646	40.5	324	2	Q7XAS1 ARATH
7	497	12.2	798	2	Q8YTL0 ANASP
8	448.5	11.0	626	2	Q7NFP3 GLOVI
9	442	10.9	631	2	Q93AK0 SYNPF
10	442	10.9	631	2	Q5N029 SYNPF
11	431	10.6	819	2	Q8VY16 ARATH
12	429	10.6	714	2	Q55559 SYNPF
13	425	10.5	841	2	Q8LUL2 ARATH
14	423	10.4	673	2	Q8DKU7 SYNEL
15	344.5	8.5	702	2	Q7V0H1 PROMP
16	340	8.4	653	2	Q7U5S7 SYNPF
17	318	7.8	685	2	Q7VAU3 PROMA
18	317	7.8	661	2	Q7V8J7 PROMM
19	155.5	3.8	5362	2	Q93155 BACSU
20	152.5	3.8	1263	2	Q5KIT6 CRYNE
21	148.5	3.7	5363	2	Q7OJ29 BACAM
22	148	3.6	5363	2	Q6YK40 BACSU
23	145	3.6	2297	2	Q8YL08 ANASP
24	143	3.5	2089	2	Q39478 9STRA
25	141.5	3.5	1329	2	Q4IA74 GIBZE
26	141	3.5	2076	2	Q6LZY0 METWP
27	140.5	3.5	1835	2	Q6BMT4 DEBHA
28	139	3.4	1027	2	Q7RYV6 NEUCR
29	138	3.4	1274	2	Q5STK0 CRYNE
30	136.5	3.4	1954	2	Q7MIT8 VIBVU
31	134	3.3	678	2	Q8PHZ9 XANAC

32	134	3.3	1238	2	Q9LEP9 BRANA	Q9LEP9 brassica na
33	133	3.3	1241	2	Q9UKK5 HUMAN	Q9UKK5 homo sapien
34	132.5	3.3	1011	2	Q9JGJ3 9RICK	Q9JGJ3 rickettsia
35	132.5	3.3	1951	2	Q8DB34 VIBVU	Q8DB34 vibrio vuln
36	132.5	3.3	2111	1	MCAS MYCBO	Q02251 mycobacteri
37	132	3.2	926	2	Q4JUH6 CORJK	Q4JUH6 corynebacte
38	132	3.2	3213	2	Q4HMX8 GIBZE	Q4HMX8 gibberella
39	131.5	3.2	1194	1	DFOL VZVD	P09252 varicella-z
40	131.5	3.2	1194	2	Q6QCM7 HHV3	Q6QCM7 human herpe
41	131.5	3.2	2111	2	P96291 MYCTU	P96291 mycobacteri
42	131	3.2	852	2	Q6BSS1 DEBHA	Q6BSS1 debaryomyce
43	130.5	3.2	876	1	DP01 BACST	P52026 bacillus st
44	130.5	3.2	1009	2	Q93G19 RICCN	Q93G19 rickettsia
45	130.5	3.2	1009	2	Q93GJ2 RICCN	Q93GJ2 rickettsia

ALIGNMENTS

RESULT 1

Q7XAS0 ARATH
ID Q7XAS0 ARATH PRELIMINARY; PRT; 801 AA.
AC Q7XAS0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Division protein.
GN Name=ARC6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE
RX MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,
RA Osteryoung K.W.;
RT "ARC6 is a J-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933(2003).
DR EMBL; AY221468; AAQ18645.1; -; Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaJ N.
SQ SEQUENCE 801 AA; 88306 MW; D09E27253E46FC6 CRC64;

Query Match 100.0%; Score 4063; DB 2; Length 801;
Best Local Similarity 100.0%; Pred. No. 4.7e-250;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEALSHVGIGLSPFLCRPPATTKLRSSNTTICSASKWADRLLSDFNFTSSSSSS	60
Db	1	MEALSHVGIGLSPFLCRPPATTKLRSSNTTICSASKWADRLLSDFNFTSSSSSS	60
Qy	61	FATATTATLVSPPSDRPHVPIPIDFVVLGAOTHTLTDGIRRAFEARVSKPPQFG	120
Db	61	FATATTATLVSPPSDRPHVPIPIDFVVLGAOTHTLTDGIRRAFEARVSKPPQFG	120
Qy	121	FSDDALISRRQILQAACETLSNPRSRREYNEGLLDDDEATVITDVPMDKVPALCVLQEG	180
Db	121	FSDDALISRRQILQAACETLSNPRSRREYNEGLLDDDEATVITDVPMDKVPALCVLQEG	180
Qy	181	GETEIVLRVGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDPPDFTGYEFVEFAL	240
Db	181	GETEIVLRVGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDPPDFTGYEFVEFAL	240
Qy	241	KLLOEGGASLAPDLRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS	300
Db	241	KLLOEGGASLAPDLRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS	300

```
QY 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
Db 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
QY 361 FIGKKPHLLQDADKQFOOQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
Db 361 FIGKKPHLLQDADKQFOOQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGKLLLETWLAGVVPFRDRTDK 480
Db 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGKLLLETWLAGVVPFRDRTDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATMARI GAHVKASAMQALQKVPFSRYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATMARI GAHVKASAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVNGRDEPGVFTAEAVRPSNPENFTNDYAIRAGVSSSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVNGRDEPGVFTAEAVRPSNPENFTNDYAIRAGVSSSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
QY 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
Db 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVVDYTLILKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVVDYTLILKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLVAS 801
Db 781 RYEVFWSKSGWKITEGSLVAS 801

RESULT 2
Q7XAR9 ARATH
ID Q7XAR9 ARATH PRELIMINARY; PRT; 801 AA.
AC Q7XAR9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Division protein.
GN Name=ARC6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]_TaxID=3702;
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,
RA Osteryoung K.W.;
RT "ARC6 is a J-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933(2003).
DR EMBL; AY221469; AAQ18646.1; -; mRNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaU_N.
SQ SEQUENCE 801 AA; 88248 MW; 7E2E1B3FD4BE4B61 CRC64;

Query Match 99.7%; Score 4052; DB 2; Length 801;
Best Local Similarity 99.8%; Pred. No. 2 4e-249;
Matches 799; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEALSHVGIGLSPFQLCRLPPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
|||||
Db 1 MEALSHVGIGLSPFQLCRLPPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
QY 61 FATATTATLVSPSPSIDRPERHVPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFG 120
|||||
Db 61 FATATTATLVSPSPSIDRPERHVPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDDALISRRQIIQAACETLSNPRSRREYNELGLDDEATVITDVPWKVPKALCVLQEG 180
|||||
Db 121 FSDDALISRRQIIQAACETLSNPRSRREYNELGLDDEATVITDVPWKVPKALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALFVSRDAMALDPPDFTTGYEFVEBAL 240
|||||
Db 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALFVSRDAMALDPPDFTTGYEFVEBAL 240
QY 241 KLGQEGASSLAPDLRAQIDETLEEITPRVYLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
|||||
Db 241 KLGQEGASSLAPDLRAQIDETLEEITPRVYLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
Db 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
QY 361 FIGKKPHLLQDADKQFOOQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
Db 361 FIGKKPHLLQDADKQFOOQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGKLLLETWLAGVVPFRDRTDK 480
Db 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGKLLLETWLAGVVPFRDRTDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATMARI GAHVKASAMQALQKVPFSRYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGSPPLAAATMARI GAHVKASAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVNGRDEPGVFTAEAVRPSNPENFTNDYAIRAGVSSSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVNGRDEPGVFTAEAVRPSNPENFTNDYAIRAGVSSSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
QY 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
Db 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVVDYTLILKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVVDYTLILKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLVAS 801
Db 781 RYEVFWSKSGWKITEGSLVAS 801

RESULT 3
Q9FIG9 ARATH
ID Q9FIG9 ARATH PRELIMINARY; PRT; 801 AA.
AC Q9FIG9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MDH9
DE (Hypothetical protein At5g42480).
GN Name=At5g42480;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]_TaxID=3702;
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=99156233; PubMed=10048488;
RX MEDLINE=99156233; PubMed=10048488;
```

RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RA "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT Sequence features of the regions of 1,081,958 bp covered by seventeen
RT physically assigned P1 and TAC clones.";
RA DNA Res. 5:379-391(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Huan V.W., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016888; BAB10489.1; -; Genomic_DNA.
DR EMBL; AY091075; AAM13895.1; -; mRNA.
DR EMBL; AY150490; AAM12907.1; -; mRNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaJ_N.
KW Hypothetical protein.
SQ SEQUENCE 801 AA; 88260 MW; 608E776B73FECF CRC64;

Query Match 99.7%; Score 4051; DB 2; Length 801;
Best Local Similarity 99.8%; Pred. No. 2.7e-249;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSFQLCRLPATTYKLRSHNTSTTICSAKWADRLSDNFNTSDSSSS 60
DB 1 MEALSHVIGLSFQLCRLPATTYKLRSHNTSTTICSAKWADRLSDNFNTSDSSSS 60

QY 61 FATATTATVLSPPSIDRPERHVPIDFVYVLGAQTHFTDGIKRAFEARVSKPPQFG 120
DB 61 FATATTATVLSPPSIDRPERHVPIDFVYVLGAQTHFTDGIKRAFEARVSKPPQFG 120

QY 121 FSDDALISRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPWPKVPGALCVLQEG 180
DB 121 FSDDALISRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPWPKVPGALCVLQEG 180

QY 181 GETEIVLRVEALLKERLPKSKQDVVLVMAFLDVSRDAMALDPDFITGYEFVEEAL 240
DB 181 GETEIVLRVEALLKERLPKSKQDVVLVMAFLDVSRDAMALDPDFITGYEFVEEAL 240

QY 241 KLIQEGGASLAPDLRAQIDETLEEITPRVYLELGLPLGDDYAARKLNGLSGVRNLTWS 300
DB 241 KLIQEGGASLAPDLRAQIDETLEEITPRVYLELGLPLGDDYAARKLNGLSGVRNLTWS 300

QY 301 VGGGGSALVGGITREKFMEAFRLMTAAEQVDLFVATPSNI PAESFEVVEVALVAQA 360
DB 301 VGGGGSALVGGITREKFMEAFRLMTAAEQVDLFVATPSNI PAESFEVVEVALVAQA 360

QY 361 FIGKPHLLQADKQFQOLQAQVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADKQFQOLQAQVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420

QY 421 CRMWGLDSDSQYRNPAIVEFVLENSNRDNDLPGLCLETLWLAGVVFPRDRTKDK 480
DB 421 CRMWGLDSDSQYRNPAIVEFVLENSNRDNDLPGLCLETLWLAGVVFPRDRTKDK 480

QY 481 KFKLGDDYDDPMVLSYLERVEVVOQSPLAAATMARIAGAHVKASAMQALQKVPSPRYTD 540
DB 481 KFKLGDDYDDPMVLSYLERVEVVOQSPLAAATMARIAGAHVKASAMQALQKVPSPRYTD 540

QY 541 RNSAEPKDVQETVFSVDVPGVNNVGRDGPVFTAEAVRPSSENFETNDYAIIRAGVSSSSVD 600
DB 541 RNSAEPKDVQETVFSVDVPGVNNVGRDGPVFTAEAVRPSSENFETNDYAIIRAGVSSSSVD 600

QY 601 ETTVMSVADMLKEASVKILAAAGVAIGTISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660
DB 601 ETTVMSVADMLKEASVKILAAAGVAIGTISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660

QY 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVL DGRMLKIWTDRAA 720
DB 661 GSVRADDSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVL DGRMLKIWTDRAA 720

QY 721 ETAQGLGVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQGLGVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780

QY 781 RYEVFWSKSGWKITEGSLAS 801
DB 781 RYEVFWSKSGWKITEGSLAS 801

RESULT 4
Q6Z729_ORYZA
ID Q6Z729_ORYZA PRELIMINARY; PRT; 760 AA.
AC Q6Z729_ORYZA
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DE Plastid division protein.
GN Name=P0575F10.2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
RT clone:P0575F10.2";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004895; BAD07942.1; -; Genomic_DNA.
DR Gramene; Q6Z729; -;
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaJ_N.
SQ SEQUENCE 760 AA; 84134 MW; 2C44684862795B2F CRC64;

Query Match 43.7%; Score 1775.5; DB 2; Length 760;
Best Local Similarity 47.9%; Pred. No. 2.8e-104;
Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16;

QY 12 SPFQLCRLPATTYKLRSHNTSTTIC-SASKWADRLSDNF-----TSDSSSSSFATAT 65
DB 14 APFAFSLPRPRPRPRPPPPHPSAACRAASWAERLFAFLLPTAAPSPPSPAPAPAA 73

QY 66 TTATVLSLPPSIDRPERHVPIDFVYVLGAQTHFTDGIKRAFEARVSKPPQFGSDDA 125
DB 74 APSASPPVLPDPAERSLPLQVDYFVIGAEAPHFLGDIRAFARFAIRAKPQYGYSTDA 133

QY 126 LISRRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPWPKVPGALCVLQEGETRI 185
DB 134 LVGRRQLQAHDTLMQNRSRTQYDRALSENREERALTWDIADK-----EAGEALA 184

QY 186 VLRVGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDPDFITGYEFVEEALKLQEG 245
DB 185 VLVTGEQLLDLRPPKRFKQDVVLVMAFLDVSRDAMAAASPPDVTGCCVLEALKLQEG 244

```
QY 246 EGASSLAPDLRAQIDTLEBEITPRYVLELGLPLGDDYAAKRLNGLSGVRNLSVGGGG 305
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 DGASNLAPDLLSQIDTLEBEITPRCVLELLESLPIDTEHHKQEGLOGARNILWSVGRGG 304
QY 306 ASALVGLTREKPMNEAFIRMTAAEQVDLFVATPSNIPAESFEVYVALALVAQAFIGKK 365
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 IATVGGGFSREAFMNEAFIRMTSIEQDMDFSKTPNSIPPEWFEIYNVALAHVAQAIIISK 364
QY 366 PHLLQADAFQFOLOQAQVMAMEIPAMLYDTRNNWEIDFGLRGLCALLIGKVDCEMML 425
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 PQIIMMADLLFQLOKFNIGS---HYAYDN---EMDLALERAFCSLLVGDVSKCEMML 416
QY 426 GLDSEDSQVRNPAIVFVLNSN-RDNDLPLGLCKLLETWLAGVFPFRDTRDKKFKL 484
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 GIDNESSPYRDPKILBFIVTNSISSENDLLPGLCKLLETWLI FEVEPRSDTRGMQFRL 476
QY 485 GDYDDPMVLSYLERVEVVGSGPLAAATWARIAGAE-----HVKASAMQALQKVP-SR 537
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 PRNDSYDTKYTRYEMAFSGKLGWKITEGAVLKS 760

RESULT 5
Q7PC78 Oryza
ID Q7PC78 Oryza PRELIMINARY; PRT; 760 AA.
AC Q7PC78;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Plastid division protein.
GN Name=Arc6;
OS Oryza sativa (indica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39946;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Ert H.,
RA Osteryoung K.W.;
RT "ARC6 is a J-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933 (2003).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.
DR EMBL; BK000999; DAA01472.1; -; mRNA.
DR Gramene; Q7PC78; -
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaJ N.
SQ SEQUENCE 760 AA; 84134 MW; 2C44684862795B2F CRC64;
```

```
Query Match 43.7%; Score 1775.5; DB 2; Length 760;
Best Local Similarity 47.9%; Pred. No. 2.8e-104;
Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16;

QY 12 SPQLCRLPPATTKLRSHNTTTC-SASKWADRLSLDFNF-----TSDDSSSSSFATAT 65
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14 APFAFSLPRPRPRPPHPSAACRAASRWARLFAFPHLLPTAAPSPDPSPAPAPAA 73
QY 66 TTATLVSLPPSIDRPERHVPIDIFYQVLGAQTHFLTDGIRRAFARVSKPPQPFSDDA 125
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 AFSASFVPLFPDAAERSLPLQVDYFKVLGAEPHFLGDGIRRAFARVSKPPQPGYSTDA 133
QY 126 LISRROIQAACETLNSPRSRREYNEGLLDDEBZATVITDVPWPKVPGALCVLQGGETE 185
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 LVGRROMLQTAHDTLMNQNSRTQYDRALSERREALTMDIAWDK-----EAGEALA 184
QY 186 VLRVGALLKRLPKSKODVVLVMAFLVDVSRDAMALDPPDPIITGYEFVEBALKLQE 245
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 VLVGTGQLLDRPPKPKQDVVLAMALAYVDLSDMAASPDPVIGCVLRLAKLLQE 244
QY 246 EGASSLAPDLRAQIDTLEBEITPRYVLELGLPLGDDYAAKRLNGLSGVRNLSVGGGG 305
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 DGASNLAPDLLSQIDTLEBEITPRCVLELLESLPIDTEHHKQEGLOGARNILWSVGRGG 304
QY 306 ASALVGLTREKPMNEAFIRMTAAEQVDLFVATPSNIPAESFEVYVALALVAQAFIGKK 365
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 IATVGGGFSREAFMNEAFIRMTSIEQDMDFSKTPNSIPPEWFEIYNVALAHVAQAIIISK 364
QY 366 PHLLQADAFQFOLOQAQVMAMEIPAMLYDTRNNWEIDFGLRGLCALLIGKVDCEMML 425
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 PQIIMMADLLFQLOKFNIGS---HYAYDN---EMDLALERAFCSLLVGDVSKCEMML 416
QY 426 GLDSEDSQVRNPAIVFVLNSN-RDNDLPLGLCKLLETWLAGVFPFRDTRDKKFKL 484
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 GIDNESSPYRDPKILBFIVTNSISSENDLLPGLCKLLETWLI FEVEPRSDTRGMQFRL 476
QY 485 GDYDDPMVLSYLERVEVVGSGPLAAATWARIAGAE-----HVKASAMQALQKVP-SR 537
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 GDYDDPEVLSYLERMEGGGASHLAAAAIAKLAGAQAATAALGTVKSNAIQAQFNKVPPLIE 536
QY 538 YTDNSAEKPDQOETVFSVDPVGNVGRDCEPGVFAEAVRPSNPETNDYAIRAGVSES 597
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
537 QLDRSAMENT-----KDG-PCGYL-----ENFDQ----- 559
QY 598 SVDETTVMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSSFQKDMVSSMESD- 656
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 -----ENAPADHSRNAALKIISAGALFALLAVIGAKY-----LPRKRPLSAIRSEH 605
QY 657 --VATIGSVRADDSAL-----PRMDARTAEINIVSKWQIKSLAFGPDHRIEMLEPEVL 707
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
606 GSVAVANSVDSSTDPPALDEDPVHIPMDAKLAEDIVRKWQSIKSKALGPEHSVASLQEV 665
QY 708 DGRMLKIWTDRAAETAQGLVYDYLKLSVDSVTVSADGTRALVEATLEBSACLSDLVH 767
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
666 DGNMLKVTDRAAEIERHGWFEYTLSDVTIIDSITISLDGRRATVETATIDEAGQLTDVTE 725
QY 768 PENNATDVRTYTRYEVFWSK-SGWKITEGSLVAS 801
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 PRNDSYDTKYTRYEMAFSGKLGWKITEGAVLKS 760
```

RESULT 6

```
Q7XAS1 ARATH
ID Q7XAS1 ARATH PRELIMINARY; PRT; 324 AA.
AC Q7XAS1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Truncated division protein.
GN Name=arc6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

NUCLEOTIDE SEQUENCE.

RX MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
 RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,
 RA Oosteryoung K.W.;
 RA "ARC6 is a J-domain plastid division protein and an evolutionary
 RT descendant of the cyanobacterial cell division protein Ftn2.";
 RL Plant Cell 15:1918-1933(2003).

DR EMBL; AY221467; AAQ18644.1; -; Genomic DNA.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0031072; F:heat shock protein binding; IEA.

DR GO; GO:0051082; F:unfolded protein binding; IEA.

DR GO; GO:0006457; P:protein folding; IEA.

DR InterPro; IPR001623; DnaJ_N.

SQ SEQUENCE 324 AA; 35468 MW; 0CC5C0CC469A6325 CRC64;

Query Match 40.5%; Score 1646; DB 2; Length 324;
 Best Local Similarity 100.0%; Pred. No. 1.4e-96;
 Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSPQCLRPATTKLRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60

DB 1 MEALSHVIGLSPQCLRPATTKLRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60

QY 61 FATATTATLVSPPSIDRPERHVPIDFQVILGAOHTFTDGRRAFEARVSKPPQFG 120

DB 61 FATATTATLVSPPSIDRPERHVPIDFQVILGAOHTFTDGRRAFEARVSKPPQFG 120

QY 121 FSDALISRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPWPKVPGALCVLQEG 180

DB 121 FSDALISRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPWPKVPGALCVLQEG 180

QY 181 GETEIVLRGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDDPDFITGYEFVEAL 240

DB 181 GETEIVLRGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDDPDFITGYEFVEAL 240

QY 241 KLQEGGASSALPLDRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300

DB 241 KLQEGGASSALPLDRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300

QY 301 VGGGASALVGGITREKFMNEAF 324

DB 301 VGGGASALVGGITREKFMNEAF 324

RESULT 7

Q8YTL0 ANASP PRELIMINARY; PRT; 798 AA.

AC Q8YTL0;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE All2707 protein.

GN OrderedLocusNames=all2707;

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

OX NCBI_TaxID=103690;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=21595285; PubMed=11759840;

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,

RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing

RT cyanobacterium Anabaena sp. strain PCC 7120.";

RL DNA Res. 8:205-213(2001).

DR EMBL; BA000019; BAB74406.1; -; Genomic_DNA.

DR FIC; AD2144; AD2144.

DR HSSP; P08622; 1B0Z.
 DR GO; GO:0031072; F:heat shock protein binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ_1.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS00076; DnaJ_2; 1.
 DR Chapterone; Complete proteome.
 SQ SEQUENCE 798 AA; 90124 MW; 3C57B66C221722AA CRC64;

Query Match 12.2%; Score 497; DB 2; Length 798;

Best Local Similarity 24.6%; Pred. No. 9.6e-23;

Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;

QY 84 VPIPIDFYQVLGAOHTFTDGRRAFEARVSKPPQFGFSDALISRQILQAACETLSNP 143

DB 11 VRIPLDYVRIILGLPLAASDEQLRQAYSDRIVQLPRREYSQAALSKQLIEEAYVVLSDP 70

QY 144 RSRREY-----NEGLLDDEATVITDVPWPKVPGALCV 176

DB 71 KERSYDQLYLAHAYDPDAAATTKVAENRGDSNNGHFDVQSLSI--EVSSEBELIGALLI 128

QY 177 LOEGGETEIVLRVGEALL-----KERLPKSP-----KQDVLMALAFLDVSR 219

DB 129 LOELGEYELVLKLGKRNLYGNQNGTASTRNGNHRTPPEFLDSSERPDILLTVALASLELGR 188

QY 220 D-----AMALDDPDFITGYEFVEEALKLQEGGASSLAPDLRAQIDETLEETPR 269

DB 189 EQWQCHVENAALSLE-----TGQE-----VLPSEG---IPPSVQAEIQADLYKLRY 233

QY 270 YVLELLGLPLGDDYAAKRLNGLSGVRNLIWSVGG--GGASALVGGITRE---KPMNEAF 325

DB 234 RILELLALP--QEKTIERHQGLDLQSLIDDRGGIDGTGNDQSLGNDIDFLRFIQQLRH 291

QY 326 MTAEOVDLFAVTSNIPAESFEVVEVALALVAQAFIGKKPHLLQADKQFOOLOQAKVM 385

DB 292 LTVAEQHKLPDG--ESKRPS--AVATYLVAYASTARTQORPALIRHAKQILMLRSKQ-- 347

QY 386 AMEIPAMLYDRNNWEIDFGLERGLCALLIGKVDRCMWLGDSDSQVYRNPAIVEFVLE 445

DB 348 -----DVHLEQSLCALLLGQTEEATRVLELSQE---YEALALI-----R 383

QY 446 NSNRDDNDLFLGCKLLETLAGVVFPRFRDTKDKKFLGDDYDDPMVLSYLERV----- 500

DB 384 EKSQDSPDLLPGLGLYAEQWLQNEVFPHFRLDSRQOASLKDVFANQVQVQVLEALPND 443

QY 501 -----EVVQGSPLAAATMAR--IGAEE--VKASAMQALQKVFPS---RYTD 540

DB 444 TTNEWAVINRQSPQSGNYSGGTTPVAKRPVGKARRPGEASTRPVPQRSHPSSEVNRP 503

QY 541 RNSAEPKDVQETVFSVDVPGNNVGRDGE-----PGVFIAEAVRSENFETND 587

DB 504 QNRTDPDELPELTPSHRRPSSNFTTARENISTTDAITDNYPPPEIPEVRASRP----- 555

QY 588 YAIRAGVS---ESSVDETTV-----EMSVADMKEASVKILAA 622

DB 556 --VQPGVSGYTGSTPPRQTPKRRRKKPQAVNVRGHSIHQQQPSPTLGRKTRLLWIVL 613

QY 623 GVAIGLISLF---SQKY-FLKS-----SSSFQKMDVMSSMESDVATIGSVRADDSEALPR 673

DB 614 G-SLGGILLFWLIVSTTFGLMKNVFFPAPSLQGEQLSIQISQPPLEIPKNAQIQSPEVS 672

QY 674 MDARTAEIVSKWKQIKSLAFPGPHRIEMLEVLDDGRMLKIWTDRAAETAQGLVVDYTL 733

DB 673 LTEETARKIIEINWLAATKASALGAEBKIESLEIITGSALSQWRLLAQDKADNRHREYS- 731

QY 734 LKLSVDSVTVS--ADGTRALVEATEESACLSDLVH--PENNATDVRTVTTTTRVEVFWSKG 790

DB 732 HSKVDSISKSIDDPNRASVGTATVRE---LTQFYENGQKGKSSDER-LRVYELIRQDDI 787

QY 791 WKI 793

DB 791 WKI 793

```
Db 788 WRI 790
RESULT 8
Q7NFP3 GLOVI
AC Q7NFP3_1 PRELIMINARY; PRT; 626 AA.
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Glr3481 protein.
GN OrderedLocNames=g1r3481;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Gloeobacteriales; Gloeobacter.
OX NCBI_TaxID=33072;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PCC 7421;
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids.";
RL DNA Res. 10:137-145(2003).
DR EMBL; BA000045; BAC91422.1; -; Genomic DNA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
KW Complete proteome.
SQ SEQUENCE 626 AA; 69297 MW; 372BAB82588B4629 CRC64;

Query Match
Best Local Similarity 11.0%; Score 448.5; DB 2; Length 626;
Matches 192; Conservative 112; Mismatches 282; Indels 159; Gaps 30;

QY 86 IPIDFYQVLGAQTHFLTDGIRRAFAERVSPPQFGSDDALIRROITLQAACBTLGNPRS 145
DB 9 LPLSYQILGVPOCTYEQVEPAFADRLAQAPREFSAVRSREHWLRACHTALGDPVR 68
QY 146 RREYN-EG----LLDDEATVITDVPMDKVPKGCALCVLQEGGETEIVLRVGEALLKRLPK 200
DB 69 RELYHREGQGLLDSSHAAV-----GLLFYELGEYQTLIERQREALAVADQP- 117
QY 201 SFQDVLVWALAFDVRDAMALDPPDFITGYFVEEALKLQEGASSLAPDLRAQID 260
DB 118 ----DTRLALAHQALQAEAYR--QGNLFLAHLVEEAEILR--GGDCLKP-VQOELQ 168
QY 261 ETLEBITPRVLELGLPLGDDVAAKRLNGLSGVRNLL---WSVGGGASALVGGLTRE- 316
DB 169 TLKURWRPERILQLLA-GAADPPSPORQEGWALLAALAEREGIGDQNDQ--SGLSREE 225
QY 317 --KFMNEAFIRMTAAQVDLIF---VATPSNIPAESFEVYEVVALALVAQAQFIGKPHLLQD 371
DB 226 FVQFLQYLRRLTVAAQQLFEREAARPS--PAAQ---YLAQAQLACGTEGSPQCVRR 280
QY 372 ADKQFOOQOAKVMWEIPAMLYVDRNNWEIDPGLERGLCALLIGKVDRCRMWGLDSED 431
DB 281 ARGHLIKLVQRQ-----DVNLELAVCALLGQVEEAQKNERSAEE 321
QY 432 SQYRNPAIVFEVLENSRDNDDLPGLCKLETLWLAGVPEPRDRTKDKKFLGDDYDDP 491
DB 322 Q-----ADVYI-KNLSQSDPLGLPCRYTDLWLAEEVPPGPRDRSGTYTLKAFAPH 374
QY 492 MVLSYLERVEVWQGSFLAAATMARICAEHVKASAMQALQKVPFSRYDRNSAEPKDVQE 551
DB 375 EVRAFLD-----DQPAAPAPAPAPARPARVIASGASS-----GTIEPGRL-- 413
QY 552 TVFSDVPVGNVGRDGEPGVFIABAVRPSNFETNDYAIRAGVSESSVETTVEMSVADM 611
DB : |||
: |||
```

```
Query Match      10.9%; Score 442; DB 2; Length 631;
Best Local Similarity 24.5%; Pred. No. 2.1e-19;
Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDFYVLGAQTHFLTDGIRRAPEARVSKPPQPFQFSDALISRRQILQAACETLSNPRS 145
DB 3 IPLDYRILCVGVQASADKLAESYDRNLNQSPSFSELSAQARRQLLEAAIAELSDPEQ 62
QY 146 RREYNE-----GLDDDEAVITDVPWDKVPKALCVLOEGGETEIVLRVGEALL----- 194
DB 63 RDRYDRFFQGLEAIEPSLELED--WQRI--GALLILLEGEYDRVVSQALBELLPDYDAS 119
QY 195 KERLPKSFQDVVLVWALAFLDVSRDAMALDPPDFTGYEFVEEALKILOEGA----- 248
DB 120 AEVRDQFARGDIALAIALSQOGLGECRCQ-----OGLYEQAQHFGRS 162
QY 249 -SSLA-----PDLRAQIDETLEETPRVLELGLPLGDDYAAKRLNGLSGVRNII--W 299
DB 163 QSALADHQRPELSRTLHQEGQURPYRILERLAQPLTAD--SDRQOGLLLQLQMLDDRQ 220
QY 300 SVGGGASALVGGTLREKFMNEAFLR-----MTAAEQVDLF---VATPSNIPAESFEVYE 351
DB 221 GIEGPGDGG--SGITLDNFI--MFLQIRGYLTAEQQLLFESERARPS--PAASF--F 271
QY 352 VALALVAQAFIGKPKPHLLQADKQFOOLQAQVWAMEIPAMLYDTRNNWEIDFGLERGLC 411
DB 272 ACYTLIARGFCDHQPSLIHRASLLHLKLS-----RMDVHIEQAIA 312
QY 412 ALLIGKVEDECRMWGLDSED-----SOVRNPAIVEFVLENSNRDNDLPLGCKLETWLA 467
DB 313 SLLGQPEEAELI--VQSQDEETLSQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVVFPRFRDTRDKKFKLGDYDDPMVLSYLERVEVVGSPPLAAATWARI GAHVKASAM 527
DB 361 TKVFPDRDLKERTAPLQPYFDDPDVQTYLDAIVEL-----VAPP 432
QY 528 QALQKVPFSRYTDRNSAEPKDVQSTVFSVDP-----VGNVGRDGEPPGVFAEVRPS 580
DB 397 -----PSDLMTPLPVEPLEVRSSLLAKELPTPATPG-----VAPP 432
QY 581 ENFETNDYAIRAGVSSVDTEVNSVADMLKESVKILAAQVAIGLISLFSQKYLKS 640
DB 433 PRRRDRSRPARTAKRLPLPWIGLV-----VVVLGGTGV-----WAWRS 475
QY 641 SSSFORKDMVSSMESDVATIGSVRADDSALP-----RMDARTAEINVSQWOKIKS 691
DB 476 RSN-----STPTPPPPVQVTLPEAVPAPSPAPVTVALDRAQAEIVLQWLAAKA 524
QY 692 LAFGPDRHIEMLPEVLDRMLKIWTDRAAETA--QLGLVYDYTLKLSDVSTVTSADGTR 749
DB 525 AALGPQYDRDLATVLTGEVLQWQGFSSQQAANTQLTSQFDH---KLTVDSVQLSDGDQR 581
QY 750 ALVEATLESACLSDLVHPENNAVDRT---YTRYEVFWSKSGWKITEGSVL 799
DB 582 AVVQAKVDE-----VEQYVRGQDLLETRDGLGVIRYQLVRENINWIKIASISLV 630

RESULT 10
QSN029_SYNP6 PRELIMINARY; PRT; 631 AA.
AC QSN029;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Cell division protein Ftn2 homolog.
GN OrderedLocusNames=sync2151.d;
OS Synchococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=269084;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sugita M.;
RT "Complete genome structure of the unicellular cyanobacterium Anacystis
```

```
RT nidulans 6301 (Synchococcus sp. PCC6301)";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Participates actively in the response to hyperosmotic
CC and heat shock by preventing the aggregation of stress-denatured
CC proteins and by disaggregating proteins, also in an autonomous,
CC dnaK-independent fashion. Unfolded proteins bind initially to
CC dnaJ; upon interaction with the dnaJ-bound protein, dnaK
CC hydrolyzes its bound ATP, resulting in the formation of a stable
CC complex. GrpE releases ADP from dnaK; ATP binding to dnaK triggers
CC the release of the substrate protein, thus completing the reaction
CC cycle. Several rounds of ATP-dependent interactions between dnaJ,
CC dnaK and grpE are required for fully efficient folding. Also
CC involved, together with dnaK and grpE, in the DNA replication of
CC plasmids through activation of initiation proteins (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The J domain is necessary and sufficient to stimulate dnaK
CC ATPase activity. Zinc center 1 plays an important role in the
CC autonomous, dnaK-independent chaperone activity of dnaJ. Zinc
CC center 2 is essential for interaction with dnaK and for dnaJ
CC activity (By similarity).
CC EMBL: AP008231; BAD80341.1; -; Genomic DNA.
DR GO: GO:0031072; F:heat shock protein binding; IEA.
DR GO: GO:0051082; F:unfolded protein binding; IEA.
DR GO: GO:0051301; P:cell division; IEA.
DR GO: GO:0006457; P:protein folding; IEA.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR PROSITE: PS00076; DnaJ_2; 1.
DR Cell division; Chaperone; Complete proteome; DNA replication;
KW Heat shock.
SQ SEQUENCE 631 AA; 70694 MW; 6546142E3B48DEE2 CRC64;

Query Match      10.9%; Score 442; DB 2; Length 631;
Best Local Similarity 24.5%; Pred. No. 2.1e-19;
Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDFYVLGAQTHFLTDGIRRAPEARVSKPPQPFQFSDALISRRQILQAACETLSNPRS 145
DB 3 IPLDYRILCVGVQASADKLAESYDRNLNQSPSFSELSAQARRQLLEAAIAELSDPEQ 62
QY 146 RREYNE-----GLDDDEAVITDVPWDKVPKALCVLOEGGETEIVLRVGEALL----- 194
DB 63 RDRYDRFFQGLEAIEPSLELED--WQRI--GALLILLEGEYDRVVSQALBELLPDYDAS 119
QY 195 KERLPKSFQDVVLVWALAFLDVSRDAMALDPPDFTGYEFVEEALKILOEGA----- 248
DB 120 AEVRDQFARGDIALAIALSQOGLGECRCQ-----OGLYEQAQHFGRS 162
QY 249 -SSLA-----PDLRAQIDETLEETPRVLELGLPLGDDYAAKRLNGLSGVRNII--W 299
DB 163 QSALADHQRPELSRTLHQEGQURPYRILERLAQPLTAD--SDRQOGLLLQLQMLDDRQ 220
QY 300 SVGGGASALVGGTLREKFMNEAFLR-----MTAAEQVDLF---VATPSNIPAESFEVYE 351
DB 221 GIEGPGDGG--SGITLDNFI--MFLQIRGYLTAEQQLLFESERARPS--PAASF--F 271
QY 352 VALALVAQAFIGKPKPHLLQADKQFOOLQAQVWAMEIPAMLYDTRNNWEIDFGLERGLC 411
DB 272 ACYTLIARGFCDHQPSLIHRASLLHLKLS-----RMDVHIEQAIA 312
QY 412 ALLIGKVEDECRMWGLDSED-----SOVRNPAIVEFVLENSNRDNDLPLGCKLETWLA 467
DB 313 SLLGQPEEAELI--VQSQDEETLSQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVVFPRFRDTRDKKFKLGDYDDPMVLSYLERVEVVGSPPLAAATWARI GAHVKASAM 527
DB 361 TKVFPDRDLKERTAPLQPYFDDPDVQTYLDAIVEL-----VAPP 432
QY 528 QALQKVPFSRYTDRNSAEPKDVQSTVFSVDP-----VGNVGRDGEPPGVFAEVRPS 580
DB 397 -----PSDLMTPLPVEPLEVRSSLLAKELPTPATPG-----VAPP 432
```

```
QY 581 ENFETNDYAIRAGVSESSVDETTVMNSVADMLKBSVKILAAAGVATGLISLFSQKYFLKS 640
Db 433 PRRRRDRSERPARTAKRLPLPWIGLV-----VVVLGGTGV-----WAWRS 475
QY 641 SSSFORQKDWSSMESDVATIGSVRADDSALP-----RMDARTAEINIVSKWKIIS 691
Db 476 RSN-----STPPPTPPVQVQLPEAVPAPSPVTVVALDRAQAETVLQNLWLAAGA 524
QY 692 LAFGPDHRIEMPEVLGDGRMLKIWTDRAAETA--QLGLVVDYTLKLSDVSTVTSADGTR 749
Db 525 AALGPQYDRDLATVLTGEVLTQWGFSSQQAQNTQLTSQFDH---KLTVDVSQVLSGDQR 581
QY 750 ALVEATLRESACLSLIVHPENNATDVRT---YTTRYEVFWMSGWNKITGSVL 799
Db 582 AVVOAKVDE---VEQVYRGDQLLETRDLGLVIRYQVLRENNINWIKIASISLV 630

RESULT 11
Q8VY16 ARATH PRELIMINARY; PRT; 819 AA.
AC Q8VY16;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein At3g19180.
GN Name=At3g19180;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
RA Carminci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074283; AAL66980.1; -; mRNA.
KW Hypothetical protein.
SQ SEQUENCE 819 AA; 90750 MW; 6720083F4864657E CRC64;

Query Match
Best Local Similarity 10.6%; Score 431; DB 2; Length 819;
Matches 200; Conservative 116; Mismatches 311; Indels 224; Gaps 32;

QY 56 SSSSFATATTATLVSLPPSIDRPRHVPIDFYQVLQAQTHFLTDGIRRAFEARVSK 115
Db 81 SRTSSLAATST-----TLPVTCYQLIGVSEQAEDEVKSVINLKKT 124
QY 116 PPQGFSDALISRRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPMDKVPALC 175
Db 125 DAEGYVTEAAAAQDLMDVRDKL---LFSEYAGNLKEKIAKPSLRIPAWLPGALC 181
QY 176 VLQEGGETEIVLRVBALKERLPSKQDVVLVMAFLDVSRDAMALDPPDFTGYEF 235
Db 182 LLQEVGQELVDIGRAALRLNDSKPYIHDIFLSMALAECAAKAFVKNVS--QGFEA 239
QY 236 VEEALKLOEE-GASSLAPDLRAQIDTELEETPRVLELGLPLGDDVAAREKELGSLV 294
Db 240 LARAQSLSKVTLGKLA--LTQIEESLEGAPPCTLDLGLPRTPEAERRRGAIAL 297
QY 295 RNILSVGGGASALVGGITREK-----PMNEAFLEMTAAEQVDLF-----VATP 339
Db 298 RELLRQ-----GLSVEASCOIQDWPCFLSQAISRLATEIVDLPLPDDDLAITRK 346
QY 340 SNIPAESFE-----VYEVALALVAQAFIKGPKHLLQDADKQFOQLQQAQKMWAMEIP 390
Db 347 NKKSLSHNQVRVIDPNCFCYVLLGHIAVGFSG-----KQNETINKAKTICECLI 396
```

```
QY 391 AMLYDTRNNWEIDFGLERGLCALLIGKVDECRMWGLG-----DSEDQVNRNPAIVEFVLEN 446
Db 397 A-----SEGVDLKFEAEPCFLLKQGSAAELEKLEKQLESNDSAVRNS-----ILKG 444
QY 447 SNRDDNDLPLCKLLETWLAGVVFPRFRDTK-----DKFKLGDYDDPMVL 494
Db 445 ESRST-----ATPSLEAWLMESVLANFPDTRGCSPLANFFRAEKKYPENKMGSPSIM 499
QY 495 SY-----LERVEVQGSPLAAAAATMARIGAEHVKASAMQALQKVFPSRYTRDRNSAEKPD 548
Db 500 NHKTNORPLSTTQFVNS-----QHL-----YTAVQLITFD 531
QY 549 VQTVFSVDPVGNVNGVGRDGPVFIABAVRPSNFETNDYAIRAGVSESSV-DETTVMS 607
Db 532 LQSPVVSAAK---NNDE-----TSASMPFSVLKRN-----LGVHKNKIWDEW--LS 571
QY 608 VADMLKEASVKILAAAGVAIGLISLFSQY-----636
Db 572 QSSLIGRVSVVAL-----LGCTVFFSLKLSGIRSLQSMPIVSARPHSESDSFLWKTE 626
QY 637 ---FLKSSSSFORQKDWSS-----MESDVATIGSVRADDSALPR 673
Db 627 SGNFRKULDSVNRNGIVGNIKVLIDMLKMGCEHPDLYLKSSQGSATSLSHSASELHKR 686
QY 674 -MDARTAEINIVSKWKIISLAFGPDHRIEMLPEYLDGRMLKIWTDRAAETAQGLGVY-DY 731
Db 687 PMDTEEAELVRQWENKVAEALGPTHQVYSLSEVLDESMLVQM-QTLAQTAAEAKSCYWRP 745
QY 732 TLKLSDVSTVTSAD---GTRALVEATLRESACLSLIVHPENNATDVRTTYTRTEVFWMSK 788
Db 746 VLLHLEVLQAHIFEDGTAGEAAEIALLEBAELVDSEQPK-NAKYTYSTYKIRYLKQEE 804
QY 789 SG-WKITEGVS 798
Db 805 DGLWKFQCSDI 815

RESULT 12
Q55559 SYN3 PRELIMINARY; PRT; 714 AA.
AC Q55559;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE S110169 protein.
GN OrderedLocusNames=s110169;
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima T., Hikosawa M., Sugita M., Shimamura M., Nakazaki N.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; BA000022; BAA10060.1; -; Genomic_DNA.
DR FIR; S76082; S76082.
DR HSSP; P08622; IBQZ.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0000917; P:barrier septum formation; IDA.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DnaJ 2; 1.
DR Chaperone; Complete proteome.
SQ SEQUENCE 714 AA; 79424 MW; 51B52C16F405ED3E CRC64;
```

```
Query Match      10.6%; Score 429; DB 2; Length 714;
Best Local Similarity 25.1%; Pred. No. 1.7e-18;
Matches 194; Conservative 124; Mismatches 331; Indels 124; Gaps 30;

QY 86 IPIDFYQVLGAQTHFLTDGIRRAPEARVSPQPGFSDALISRRQILOAAACETLSNPRS 145
DB 3 IPLDFYRILGIPQSGGETIEQAQYDRLLQLPRREFSDAATVLENQLLATAYETLRDEK 62
QY 146 RREYNE---GLLDBE--EATVITDVPWVKVP---GALCVQEGGETEIVLVRVGEALLKE 196
DB 63 ROAYDQEWGAMDEALGALPLATTPELECSPEQBIGALLILLDGEYELVVKYGEPLVD 122
QY 197 RLPKS---FKQDVVLWALAFDVSDDAMALDPDPFITGYEFVFEA-IKLQEBEGASSIAP 253
DB 123 PNPPAGGLPDYLLSVILAHWSLSREWQQQ-----YEFATASLAKALARLQDDNFP 176
QY 254 DLRAQIDETLEETPRVLELLGLP-LGDDVYAARKLNGLSGVNLSVGG-GGASALVG 311
DB 177 ALBAEIRQELRYLRPYRILLELLAKEGQEB---ORQGLALLQWQDRGIEGKGEDYS 233
QY 312 GLTRE---KFMNEAFLEMTAAEQVDLFVATPSNIPAESFEVFEYVALALVAQAFTGKKPHL 368
DB 234 GLGNDDFLKFIHQRLCHLTVAEQNALFL--PESQRPISLVASYLAVHSLMAEGVKEQDPA 291
QY 369 LQADKQFOQLOQAQVWAMEIPAWLYDTRNNWEIDFGLERGLCALLIGKVDCEKMWGLD 428
DB 292 IVEAKSLIIQLENCQ-----DLALEKVICELLIGQTEV--VLAAD 330
QY 429 SEDSQYRNPAIVFVLNSNRDNDLPGCLKLETLWAGVPPRPDTRDKKFKLGDY 488
DB 331 QGD-----PKIVA-GLESKLATGEDPLTAFYTFTEQWLEBEEIVPYFRDLSPETLSPKAYF 384
QY 489 DDPMVLSYLERVEVQ-----GSP-LAAATWARIAGAEHVKASAM-QALQKVPFSPRYT 539
DB 385 NNFVSQOYLEQEPDFTTNSFASPALLSTATSETPMWSSAALPDRFLTSTVPS---441
QY 540 DRNSAEPKQVQETVFSVDPVGNVGRDGEFPGVFAEAVRPSENFENDYAIRA-----592
DB 442 -RRGRSPRRSRDDVFP-----SADNSSGLAVT-TLSPAIAYDTHSLGTNGIGGDSTS 491
QY 593 -GVSESVDETTVE-MSVADMLKEASVK-----ILANGVALIGLSFSQK-----635
DB 492 NGFSSNAPESTSKHSPRRKRKRVTIKPVRFGLICLAGIVGGATALINRTGDPILG 551
QY 636 -----YFLKSSSSFKQDMVSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWQK 688
DB 552 LLEDPLDVFL-----DQSEFIPDEATSNLILSQ-----PNFNQQVGQVWVQGWLD 598
QY 689 IKSLAFPGDHRHLEMLPEVLDGRMLKIWTDRAETAQLGLVYDYLTLKLSVDSTVS-ADG 747
DB 599 SKKLAFQNYDVGLAQSLVAPNLQAQGR-AQRDQAQKYVHYEHKQLIAYQVNPQDP 657
QY 748 TRALVATELSEACSLDLVHPENNATDVR--TYTTRVEVFWKSGWKITEGSVL 799
DB 658 NRATVTARVEEISQPTLGNQQQGSATKDDLTVRYQLVHQGWKIDQIQV 710

RESULT 13
Q9LJL2_ARATH PRELIMINARY; PRT; 841 AA.
ID Q9LJL2;
AC Q9LJL2;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone: MW111.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC Rosids; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
EX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones.";
RN DNA Res. 7:217-221(2000).
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP000419; BAB02958.1; -; Genomic DNA.
SQ SEQUENCE 841 AA; 93216 MW; 7DBEE72618EB8B97 CRC64;

Query Match      10.5%; Score 425; DB 2; Length 841;
Best Local Similarity 23.0%; Pred. No. 4e-18;
Matches 200; Conservative 122; Mismatches 308; Indels 240; Gaps 35;

QY 56 SSSSPATATTATLVSIPPSIDPRHVPIPIDFYQVLGAQTHFLTDGIRRAFEARVSK 115
DB 81 SRTSSLAASTST-----IELPVTCTYQIGVSEQAQAEKDEVKSVINLKKY 124
QY 116 PPOFGFSDDALISRRQILOAAACETLSNPRSREYNEGLLDDDEEATVITDVPWPKVPGALC 175
DB 125 DABEGYTMEAAAARQDILLMDVRDKL---LFSEYAGNLKEKIAPKSPLRIPWAWLPGALC 181
QY 176 VLOEGGTEIVLVRVGEALLKERLPKQKQDVVLWALAFDVSDDAMALDPDPFITGYEF 235
DB 182 LLOEVGQEKLVLDIGRAALRNLDKPYTHDIFLSMALAECIAKAAFEVNVKVS--QGPEA 239
QY 236 VEEALKILOEB-GASSIAPDLRAQIDETLEITPRYVLELLGLPLGDDYAAKRLNGLSGV 294
DB 240 LARAQSPFKSKVTGKLA--LLTQIEESLELAPPCTLDLGLPRTPENAEIRGAIAAL 297
QY 295 RNILWSVGGGASALVGLTREK-----FMNEAFLEMTAAEQVDLF-----VATP 339
DB 298 RELRQ-----GLSVEASQIQDWCFILSQALISRLLAIEIVLLPWDDLIATRK 346
QY 340 SNTPAESFE-----VYEVALVALVAQAFIGKKPHLLQDADKQFOOLOQAQVWAMEIP 390
DB 347 NKKSLESHNRQVWIDFNCFYWVLLGHIAVGSF-----KQNETINKAKTICECLI 396
QY 391 AMLYDTRNNWEIDFGLERGLCALLIGKVDCE--RMWLGLDSE-----DSQY 434
DB 397 A-----SEGVDLKFEAFCSFLKQLSATGPTCWTIAQGSAAEAEKLEKQLESNDSAV 449
QY 435 RNPAIVFVLNSNRDNDLPGCLKLETLWAGVPPRPDTRK-----DKKF 482
DB 450 RNS-----ILKESRSTG-----ATPSLEAWMESVLANFPDTRGCSFSLANFFRAEKY 499
QY 483 KLGDIYDDPMVLSY-----LERVEVVGSPALAAATWARIAGAEHVKASAMQALQKVPFS 536
DB 500 PENKMGSPSIMNHKTNRPLSTTQFVNSS-----QHL-----532
QY 537 RYTDNRSAEPKQVQETVFSVDPVGNVGRDGEFPGVFAEAVRPSENFENDYAIRAGVSE 596
DB 533 -YTAVEQLTPTDLQSPVVSAK--NNDE-----TSAMSPSVQLKRN-----LGVHK 574
QY 597 SSV-DETTVMSVADMLKEASVKIL-----AAGVAIGLISLF-----SQ 634
DB 575 NKIWDW-----LSQSSLLGRVSVALLGCTVFFSLKLSGIRSGRLQSPISVSARPHSESD 631
QY 635 KYFLKSSSSFKQDMVSSMESDVATIGSVRA-----665
DB 632 SFLKWTESGNFRKNLDSVNRNGI--VGNKIVLIDMLXMGHCEHPDALYLKSSGQSATSLS 689
QY 666 -DPSEALPR-MDARTAEINIVSKWQIKSLAFGPDHRIEMLEPEVLDGRM-----LK 713
DB 690 HSASELHKRPMDTEAEELRQWENKAEALGPTHQVYVLSSEVLDESMLAVQVESIFCLM 749
QY 714 INTDRAETAQLGLVY--DYTLLKLSVDSVTVSAD---GTRALVEATLEESACLSLVIPE 769
DB 750 QW-QTLTAQTAEAKSCYWRFFVLLHLEVLQAHIFEDGIAEAAIEALLLEAAELVDESQPK 808
```

```

QY 770 NNATDVRTYTRYEVFWSSKSG-WKITEGVS 798
Db 809 -NAKYSTYKIRYILKKQEDGLWKFCQSDI 837

RESULT 14
Q8DKU7 SYNEL
ID Q8DKU7_SYNEL PRELIMINARY; PRT; 673 AA.
AC Q8DKU7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
DE Tlr0758 protein.
GN OrderedLocNames=tlr0758;
OS Synchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=32046;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
CC -!- FUNCTION: Participates actively in the response to hyperosmotic
CC and heat shock by preventing the aggregation of stress-denatured
CC proteins and by disaggregating proteins, also in an autonomous,
CC dnaJ-independent fashion. Unfolded proteins bind initially to
CC dnaJ; upon interaction with the dnaJ-bound protein, dnaK
CC hydrolyzes its bound ATP, resulting in the formation of a stable
CC complex. GrpE releases ADP from dnaK; ATP binding to dnaK triggers
CC the release of the substrate protein, thus completing the reaction
CC cycle. Several rounds of ATP-dependent interactions between dnaJ,
CC dnaK and grpE are required for fully efficient folding. Also
CC involved, together with dnaK and grpE, in the DNA replication of
CC plasmids through activation of initiation proteins (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- DOMAIN: The J domain is necessary and sufficient to stimulate dnaK
CC ATPase activity. Zinc center 1 plays an important role in the
CC autonomous, dnaK-independent chaperone activity of dnaJ. Zinc
CC center 2 is essential for interaction with dnaK and for dnaJ
CC activity (By similarity).
CC EMBL; BA000039; BAC08309.1; -; Genomic DNA.
CC GO; GO:0031072; F:heat shock protein binding; IEA.
CC GO; GO:0051082; F:unfolded protein binding; IEA.
CC GO; GO:0006260; P:DNA replication; IEA.
CC GO; GO:0006457; P:protein folding; IEA.
CC GO; GO:0006986; P:response to unfolded protein; IEA.
CC InterPro; IPR001623; DnaJ_N.
CC Pfam; PF00226; DnaJ; 1.
CC SMART; SM00271; DnaJ; 1.
CC PROSITE; PS50076; DnaJ_2; 1.
KW Chaperone; Complete proteome; DNA replication; Heat shock.
SQ SEQUENCE 673 AA; 75742 MW; 3427C6B46B8CB3A CRC64;

Query Match 10.4%; Score 423; DB 2; Length 673;
Best Local Similarity 25.4%; Pred. No. 3.8e-18;
Matches 191; Conservative 105; Mismatches 307; Indels 150; Gaps 28;

QY 86 IPIDFYQLGAQTHFTDGIRRAFEARVSKPPQFGSDALISRRQILQAACETLGNPRS 145
Db 3 IPIDYYQLGLVPTQAQTEQIEQAQFRDLQLPQTHQSPPTVATRRRELIEQAVAVLREPEQ 62
QY 146 BREYNE-----GILLDEEATVITDPWDKVPGLCVLQEGGETEIVLVRGEALL 194
Db 63 RDAYDRHCRVTDPDDLIAQLDPAATTPHIEISDEQLSGALLLYELGNVAQVNVILGDAFL 122

```

```

QY 195 KERLPKSFQ-----DVIWMAFLDVSRDAMALDPPDFITGYEFVFEALKLL 243
Db 123 KQDV---FERNRPYTPAAVADITLTVALAYLGRHEW--QRQSYESAASQLEAGLQVL 177
QY 244 QEBGASSLAPDLRAQIDETLEEITPRYVLELLGLPGDDVAAKRLNGLSGVNRNLTWSYGG 303
Db 178 QR---VNLPELOEQQTENLRIPYRIELLELLALPLSD--SANRQGIILLRQMLSERGG 232
QY 304 -GGASALVGLLTRE---KFMNEAFLRMTAAEQVDLFWATPSNTPAESFEVVEYVFEALVAQ 359
Db 233 IEGRGDDRSGLTVEDFLKFTILQRLSHLTVAEQQELF-----ERESRRPSAVATYLAHV 285
QY 360 AFIGKXPHLLD-----ADKQFOOLOQAQVMMEIPAMLYDTRNNWIDFGLERGLCALL 414
Db 286 ALVARGVHELQPSYICRAKDLLOQL-----LPHQ-----DVIYLELASCILL 326
QY 415 IGKVDSCRMWGLGDSQVNRPAIVEFVLENSNRDNDLPLGLCKLLETWLAGVVRPRF 474
Db 327 LGQPTB-----ALAALDHSODQPTL--DFIRRHAG--EAGDRLPLGLYYTTQWLTEIYPAP 379
QY 475 RDTKDKKFKLGDYYDDPMVLSYLERVEVVGSGSFLAAATMARIGAELHVKASAMQALQKVF 534
Db 380 RDLGETPVALEAYFADANVQTYLEALSSEDSIAPEPPATTA----- 419
QY 535 PSRYTDRNSAEPKDVQETVPSVDPGVNNGRDSGEPGVFIAEAVRPSNFETNDYAIRAGV 594
Db 420 -----SALPEVIRPTV-AVPP-----PLSFTAETL-PLQDQSLGQGLSASA 459
QY 595 SESSVDTEVVMSVADMLKEASVK-----ILAAQVA-IGLISLPSQKQFLKSSS 642
Db 460 FTSPATATGTMPPQSPFRKRSRPNRCAQKQRTWFMGAGVLVGLGAL--AKYVWPAPT 517
QY 643 SFORKDMVSMESDVATIGSVRADDSSEAL---PRMDARTAEINIVSKWQIKSLAFGPDHR 699
Db 518 AEAPPPVTPATPVPATPTPTPQTTILAITLTPM-----ARDRLHTWQIKQAALGRPFE 573
QY 700 IEMPEVLGRMLKIWTDRAAETAQGLVVDYDITLLKLSVDSVTVS--ADGTRALVE----A 754
Db 574 VDKLTTLIAPELSRWRSAQGLKSGSYWVYTLKNLEKVEVLRQSDRVEVLAEVNEDA 633
QY 755 TLEESACLSDLVHPENNATDVRTYTRYEVFWS 787
Db 634 RFYEQGT-----RTDI--SYSDPYRVIYT 656

RESULT 15
Q7V0H1_PROMP
ID Q7V0H1_PROMP PRELIMINARY; PRT; 702 AA.
AC Q7V0H1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=PMW1287;
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=59919;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
RA Rocap G., Larimer P.W., Lamerdin J.E., Malfatti S., Chain P.,
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hegg W.R.,
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regalla W., Shah M.,
RA Shaw S.L., Stedlich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Ziesler E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
RL EMBL; BX572093; CAE19746.1; -; Genomic DNA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:39:46 ; Search time 44 Seconds
(without alignments)
1751.582 Million cell updates/sec

Title: US-10-600-070B-2
Perfect score: 4063
Sequence: 1 MEALSHVGIGLSPFQLRLP.....YEVFWKSGWKITEGSLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: Pirl:.*
2: Pirl2:.*
3: Pirl3:.*
4: Pirl4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	497	12.2	798	AD2144	hypothetical prote
2	429	10.6	714	S7082	hypothetical prote
3	145	3.6	2297	AB2494	hypothetical prote
4	143	3.5	2089	A48757	acetyl-CoA carboxy
5	139.5	3.4	879	JC4286	DNA-directed DNA p
6	138	3.4	2021	A84771	similar to ch-TOG
7	131.5	3.2	1194	DJBE28	DNA-directed DNA p
8	131.5	3.2	2111	A70668	myceroate synth
9	130.5	3.2	1026	C97783	cell surface antig
10	129.5	3.2	2110	B41110	myceroate synth
11	129	3.2	1275	T49362	hypothetical prote
12	128.5	3.2	836	T30312	pilin biosynthetic
13	128.5	3.2	876	S70368	DNA polymerase I -
14	127.5	3.1	2472	E83594	still frameshift p
15	126.5	3.1	1018	T30853	antigenic heat-sta
16	125.5	3.1	718	A81122	hypothetical prote
17	125.5	3.1	4385	T29042	hypothetical prote
18	123.5	3.0	673	C83080	probable chemotaxi
19	123.5	3.0	848	C70834	probable endoptei
20	123	3.0	3421	WZBBE6	367K tegument prot
21	122.5	3.0	899	H87513	hypothetical prote
22	122.5	3.0	2756	T30183	hypothetical prote
23	121	3.0	1381	AF2010	regulatory protein
24	120	3.0	1252	D72122	RNA polymerase bet
25	120	3.0	1262	F81548	DNA-directed RNA p
26	120	3.0	3938	T42761	Bassoon protein -
27	119.5	2.9	692	G87415	glycyl-tRNA synthet
28	119.5	2.9	698	A82593	hypothetical prote
29	119.5	2.9	831	S26675	DNA-directed DNA p

RESULT 1
AD2144
hypothetical protein all2707 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AD2144
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2144
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-798 <KUR>
A;Cross-references: UNIPROT:Q8YTL0; UNIPARC:UPI00000CE525; GB:BA000019; PIDN:BA074406.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2707
C;Superfamily: Nostoc sp. hypothetical protein all2707; dnaJ amino-terminal homology

Query Match	12.2%	Score	497	DB	2	Length	798
Best Local Similarity	24.6%	Pred. No.	3.9e-24				
Matches	207	Conservative	139	Mismatches	301	Indels	196
Gaps	33						
Qy	84	VPIPDYQVIGAGTHFLTDGIRAFARVSKPQPGFSDDALISRRQILQAACETLSNP	143				
Db	11	VRIPLDYRILGLPLAASDEQLRQAYSQAIASRKQIEAYVVLSDP	70				
Qy	144	RSREY-----NEGLDDEEATVITDPWDKVGALCV	176				
Db	71	KERSVDQLYLAAHYDPAATTKVAENRGDSNNGHFDVQSLSI--EVSSEELGALLI	128				
Qy	177	LQEGGETEIVLRVGEALL-----KERLPKSF-----KQDVVLVMAALFDVSR	219				
Db	129	LQELGEVELVLGRVNLGNQNGTASTNGNHRTPPEFLDSSERPDIILLTVALASLELR	188				
Qy	220	D-----AMALDPPDFITGVFEVEALKLOEGGASSLAPDLRAQIDETUEEITPR	269				
Db	189	EQWQOQHYENAALSLE-----TQGE-----VLFSEG---IFPSVQAEIQADLYKLRPY	233				
Qy	270	YVLEILGLPLGDDYAAKRLNGLSGVRNILSVGG--GCASALVGGITRE---KFNWEAFRL	325				
Db	234	RILELLALP--QEKTIERHQGLDLLQSLDDRGIDGTGNDQSGNLIDDFLEFQQLRRH	291				
Qy	326	MTAAEQVDLVFATPSNIPAESFEYVEVALVAQAFIGKPKPHLLQDADKQFQOQQAQVM	385				
Db	292	LTVAEQHKLFDG--ESKKEPS--AVATYLAAYASIARGFTQRPALIRHAKQILMRLSKRO--	347				
Qy	386	AMBIAPMLYDTRNNWEIDFGLERGLCALLIGKVDCECRMWLGDSQSYRNPALIVEFVLE	445				

ALIGNMENTS

Db 906 --LRGLLDNKNLVM---QTLTTIGVAAAMGPAVEKASKGILSDVLKCLGDNKKHMR 959
Qy 689 IKSLA-----FGPHRTEMPEVL-----DGRM-----LKIWTDRAAETAQLGLVYDYT 732
Db 960 ECTLAALDLMLGAVHLMDPIYIIALTIDGKMGAEGRKOLFDML-----TKQLTGLSDFV 1014
Qy 733 ----LLKLSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVR 776
Db 1015 DAHLLKPASTAMTKSADVRKAAGCISEILRVSGQEMIEKNLKDIO 1062

RESULT 7
DJB28
DNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 3
C:Species: human herpesvirus 3, varicella-zoster virus
C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 09-Jul-2004
C:Accession: B27214
R:Davidson, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A:Title: The complete DNA sequence of varicella-zoster virus.
A:Reference number: A27345; MUID:86306657; PMID:3018124
A:Accession: B27214
A:Molecule type: DNA
A:Residues: 1-1194 <DAV>
A:Cross-references: UNIPROT:P09252; UNIPARC:UPI0000129829; EMBL:X04370; NID:959989; PIDN
C:Genetics:
C:Superfamily: DNA polymerase
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 3.2%; Score 131.5; DB 1; Length 1194;
Best Local Similarity 19.0%; Pred. No. 3.8;
Matches 168; Conservative 122; Mismatches 297; Indels 297; Gaps 44;

Qy 48 SDNFETSDSSS---SSPATATTATVLSLPPSDRPERHV--PIPIDFYOVLCQAQ---TH 99
Db 205 NDATLNGDKNAFTGTSFKSPSPFRVEV---IERTDVIYDTPQCFYFVRVSPSKFTN 261

Qy 100 FLTDGIR---RAFEARVSKPPQPFQFSDALIS-----RRQI----- 132
Db 262 YLCDNFHPELKKYEGRVDAITRFLMDNPGFVGWYQLKPGVDGVRVVRPASPQLTSLD 321

Qy 133 LQAC-----ETLSNPRSRREYNGLLDDR-----EATVITDVPWVKVPGALC 175
Db 322 VEIDCMSDNLQAIPIPNDDSPDYKLLCFDIECKSGSNELAFAPDATHLEDL-----VQIQC 377

Qy 176 VL-----QEGGETIVLVGEALKERLPKSFQ-----DVLVWALAFLDVSRD 220
Db 378 LLYSIPRQSLHILLFSLGCDLPQRYVQEMKDAGLPEPTVLFFDSEFELLIAFMTLVKQ 437

Qy 221 AMALDPPDFTGYEFVEEALKLQEBGASSLAPDLRAQIDETLEEITPRVLELLGLPLG 280
Db 438 Y-----APEFATGNIYNFDWAFI-----MEKLSIYSLKL----- 468

Qy 281 DYAARKLNGLSGVRNILMSVGGGASALVGGITREKFMNEAFRLMTAAEQVDLFPATPS 340
Db 469 DVGSGINRGLFKI-----WDVGKSG-----FQRRSKVKINGLISLDWALATE 512

Qy 341 NIPAESFEVEVA-----LALVAQAFIG 363
Db 513 KLSLSYKLDSDVAREALNESKRDLPIKYDIPGYVAGNTRGITGYCIQDSALVGLKFFK 572

Qy 364 KPHLLQDADKQFQQLQAQKAWAMEIPAMLYDRNNWEIDFGLERGLCALLICKVDECRM 423
Db 573 YLPHLELSA-----VARLARITUTKAIYD-----GQVRIYTCLLGLASS--- 612

Qy 424 WLGLDSEDSQYRNPAIVEF-----VLNSNRDNDLDPGLCKLLETWLAGV-VFPR 473
Db 613 -RGFILPDGCV--PATFEYKDVIPDVGVDREEMDEDSVPTGSSGRNVGYKGVARVD- 668

Qy 474 PRDTKDKKFLGDDYDDPMVLSVLERV-EVVGQSPPLAAATWARI GAHVKASAMQALQK 532
Db 669 -PDT-----GFYIDPVVVLDFASLYPSIIQAHLN--CFTTILTNFETVK-----R 710

Qy 533 VPPSRYTDRNSAEPKDVQETVFSVDP-----VGNVVRGDGEGPCVFAE--AVRPSNPFT 585
Db 711 LNPSDY-----ATFTVGGKELFFVRSNV-RESLLGVLLKDWLAMRK----- 750

Qy 586 NDYAIRAGVSSSDVETTVEMSVADMKEASVKILA-----AGVAIGLIS----- 630
Db 751 ---AIRARIPGSSDE-----AVLLDKQQAIAKVVCVSNVYGTGVAQGFPLCLYVAATVT 802

Qy 631 -----LFSQKYLKSSSFQR-----KDMVSSMESDVA-TIGSVRRADDSSEALPRMDAR 677
Db 803 TIGRQMLLSTRDYIHNWAAFERFITAPFDIESSVLSQKAYEVKVIYGTDTDSVFIRFKGV 862

Qy 678 TAENIVSKWQK-----IKSLAFGPHRIEMLEPEVLIDGRMLKIWTDRAAETAQLGLVYDYTL 733
Db 863 SVEGIAGIKERWAHIIISTALFCPPKLECEKTFI--KLLLI-----TKKKYIGVIYGGKV 915

Qy 734 LKLSVDSVTVS-----ADGTRALVEATL-----EESACLS 763
Db 916 LMKGVLDLVRKNKCOFINDYARKLVLELLYDDTVSRAAAEASCVS 959

RESULT 8
A70668
mycocerosate synthase (EC 2.3.1.111) - Mycobacterium tuberculosis
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: A70668
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70668
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2111 <COL>
A:Cross-references: UNIPROT:P66291; UNIPARC:UPI0000036BA0; GB:Z83858; GB:AL123456; NID:g
A:Experimental source: strain H37Rv
C:Genetics:
C:Gene: mas
C:Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: acyltransferase; carrier protein; coenzyme A; phosphotransferase
F;27-426/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;534-815/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1444-1733/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;1765-1945/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;2016-2096/Domain: acyl carrier protein homology <ACP1>

Query Match 3.2%; Score 131.5; DB 2; Length 2111;
Best Local Similarity 19.9%; Pred. No. 9.4;
Matches 151; Conservative 82; Mismatches 237; Indels 289; Gaps 33;

Qy 151 EGLDDDEEATVITDVP---WDK-----VPCALCVLOEGGETEIVLRVGEALLKERL 198
Db 29 ESLRGDD--LVTEIPDRWDADDYDPEPGVGR-SVSRWGG----- 68

Qy 199 PKSFQDVLVLMALAFLDVSRDAMALDPPDFI---TGFEFVEEALKLQEBGASSLAPDL 255
Db 69 ---FLDDVAGDAEFFGISEREATSIDPQQRLLLETSWEAIEHA----- 109

Qy 256 RAQIDETLEITPRVLELGLPLGDDYAAKRLNGLSGVRNILMSVGGGASALVGGITR 315
Db 110 -----GLDPASL-----AGSSTAVFTGLTH 129

Qy 316 EKFMNEAFRLMTAAEQVDLFPATPSNI-----PAESFEVY----- 350
Db 130 EDYL-----VLTTTAGGLASPVVTVTGLNNSVASGRIHTLGLHGPANTFDTACSSGLMAVHL 186

Qy 351 -----EVALALVAQAFITGKPKHLL-----QDADKQFQLOQAK 383

Db 110 -----GLDPASL-----AGSTAVFTGLTH 129
Qy 316 EKFWNEAFRLMTAAQVDLFVATPSNI-----PAESFEVY----- 350
Db 130 EDYL---VLATTTAGGLASPYVVTGLNNSVASGRIATHTLGLHGPAMTPTDACSGLMAVHL 186
Qy 351 -----EVALALVAQAFIGKPKHL-----ODADKQFOLOQAK 383
Db 187 ACRSLHGDRLDALAGGCAVLELHACVAASAOQMLSGTGRCHSFDADAGFVRSEGCAM 246
Qy 384 VMAMEIPAMLYDTRNNNEIDFGLERGLCALLIGKVDRCMWLGDSDSOVRNPAIVEFV 443
Db 247 VLLKRLPDALRDGNR---IPAVRGATATQDGTET---LTWPSDAQV---AVTRAA 295
Qy 444 LENSNRDNDLPLGLCKLKTWLAGVYFPRDTRKFKLGDYDDPMVLSYLERVEV 503
Db 296 LAAAG-----VQPETGVV-----EAHGTGTPIGDPPIEYRSLARV-YG 332
Qy 504 QGSPAAAAATMARIGAHHVKAAMQALQKVF-----PSRYTDRNSAPKQVQETVFS 555
Db 333 AGTPCALGSAKSNMGHSTASAGTVGLIKAILSLRHGVVPPLLHFNRLPDELSDV----- 386
Qy 556 VDPGVNNGRDRGPGVFIAEAVPSENFETNDYA-IRAGVSESSVDVETTVMSVADMLKE 614
Db 387 -----ETGLFVPOQAVTPWPN--GNDHTPRKRVAVSSFGMSGTNVHAIIVEAPAE 432
Qy 615 ASYKILAAAGVA-ICLISLFSQKFLKSSSSSFQKDMVSSMESDVATIGSVRADDSEALPR 673
Db 433 ASAPESPGDAEVG-----PRUFMLSTST-----SDAL-R 461
Qy 674 MDARTAEINIVSKWK-----IKSLAFGPDHR-----IEMLPVLDGRML 712
Db 462 QTAARLATWEEHQDCVAASDLATYTLARGRRHPRVTAVVAANLPELVEGLREVADGDAL 521
Qy 713 -----KIW--TDRAAEATQIGL-----VYDYTLKLKLSVDSVTVSADGTPALV 752
Db 522 YDAVGHGDRGPMVWFSGGQSQAAMGTQLLASEPFAATIAKL---EPVIAAESGFSVT 578
Qy 753 EA-TLESACLSDLVHPENNATDVRTVTRYEVFWSKSG 790
Db 579 EAITAQOVTGIDKQVPAVFAVQVALAATMEQTYGVYRPG 617

RESULT 11
T49362
hypothetical protein B1D1.160 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
R;Schulte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49362
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1275 <SCH>
A;Cross-references: UNIPROT:Q9P6A4; UNIPARC:UPI000006B2D3; EMBL:AL355927; GSPDB:GN01116;
A;Experimental source: BAC clone B1D1; strain OR74A
C;Genetics:
A;Gene: NCSP:B1D1.160
A;Map position: 6
A;Introns: 24/3

Query Match 3.2%; Score 129; DB 2; Length 1275;
Best Local Similarity 20.9%; Pred. No. 6.2;
Matches 114; Conservative 84; Mismatches 210; Indels 138; Gaps 26;
Qy 325 RMTAARQVDLFVATPSNIPAESFEVVEVALVAQAFIGKPKHL----- 368
Db 39 RNTPATPTAYFAALLSLNNENIATPVVLLDVTVPF-APKPILOAKFTQITILAPVLS 97
Qy 369 LQADKQFOLOQAKVMAMEIPAMLYDTRNNWEI---DFGLERGLCALLIGKVDCEMML 425

Db 98 LPDAD---APLMRASIGSL-LLLAQDATQWELGTAIGPRRAVAGLLA-----L 144
Qy 426 GLDSEDS-QYRNPAIVEFVLENSNRDNDLPLGLCKLJLTLWLAGVYFPRDTRK----- 479
Db 145 ALDPRPKVRKAQALRNILKNPPSPSLDHPAAPMCAETAMQSL-----RLVADKAAQ 199
Qy 480 -KKFKLGDYDDPMVLSYLERVEVQ-----GSPAAAAATM-----ARIGAHHVKAAM 527
Db 200 RKEKTTDSTHDPDLIHALQLVKAVASGGGWPFSKSIESLCELLLSIARTGNHMSMAVF 259
Qy 528 QALQKVPSPRYTDR-NSAEPKDVQSTVFSVDPVGNV-----VGRDGPVGFIAEA 576
Db 260 EIFEMFEFGMADEVASAKLPRLLLEITIRELPAPNDTQLLPWIAILSRAYD-----VAQ 315
Qy 577 VRPSENF-ETNDYAIRAGVSES-----SVDETTVMSVADMLKEASV-----KIL 620
Db 316 ISPAETQELVDPTFLVAGYLESQHKNIRVASECLVSLANCVPKEAILEPSIDPKVI 375
Qy 621 AAGVAI--GLISLFSQKFLK-----SSSFQKDMVSSMESDVATIGSVRADDSE 669
Db 376 QQLVKVVEGLLTVOYQAAMMETFNVLGAMFADFRAWQANPYLLSV---VKSIGEIRGDSF 432
Qy 670 ALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLP-----EVLDCRMLKLTWDRAAETAOL 725
Db 433 A-----GKQEADEVILGKAIR-----AMGPEAVLNVLPNLAKPVKQGPGRW---ML 476
Qy 726 GLVYDVT---LLKLSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRTYTR 781
Db 477 PLRDYTSNTNLAHFKSELVPLSALMFORVIE-----HQEKNHIKIFETV 523
Qy 782 YEYFWS 787
Db 524 VQIQWS 529

RESULT 12
T30312
pilin biosynthetic protein - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Mar-2000
C;Accession: T30312
R;Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Mattick, J.S.
submitted to the EMBL Data Library, November 1996
A;Description: Pseudomonas aeruginosa chemotactic transduction genes pilA, chpA, chpB and
A;Reference number: Z20819
A;Accession: T30312
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-836 <WHI>
A;Cross-references: UNIPARC:UPI00001792CF; EMBL:U79580; NID:G3241967; PID:G3241968; PIDN:
C;Genetics:
A;Gene: pilA
C;Superfamily: Pseudomonas aeruginosa pilin biosynthetic protein

Query Match 3.2%; Score 128.5; DB 2; Length 836;
Best Local Similarity 20.0%; Pred. No. 3.4;
Matches 158; Conservative 106; Mismatches 292; Indels 235; Gaps 35;
Qy 124 DALISRRQILQAAACETLSNPSRREYNEGLLDDBEATVITDVPWMDKVPFALCVLQEGGE- 182
Db 24 ETLKQARQALEAFVENPQDPTMR-----FCLTYV--HQVQGTLMVVEFYGAA 69
Qy 183 --TEIVLVRGCEALLKERLPKSFQDQVVLVMALEFLVSRDAMALDPPDF-----ITGYE 234
Db 70 LLAERMEQLVQALLDGRVNPQGEALEVLMQAILQLPVYLDRIQTARRDLPMVLPFLNDL 129
Qy 235 FVEALKLLOEGGASLAPDL---RAQID-----ETLEETTPRVLELLGLP 278
Db 130 RAARGEKLLSE--TSLFAPDLSQRPQDGEAIAQLRDELGLGLLRKLRQTOQMALVGLL 187
Qy 279 LGDDYAAK-----RLNGLSGVRNI--LWSVGGGGSALVGLTRKTFMNEAFLRMT 327

Db 188 RQDVATSLGYLARVYARLEGICREAPLGLWSI-----ASGLVEGLANGSVVNSASVR-T 242
Qy 328 AAEQVD-----LFVATPSNI-----PAESFEYVEVALALVAQAFIG 363
Db 243 LLRLDRELKRLVEQAGDGLNQAAPDELKVNLLFYVAKAPSQSPRILAL----- 291
Qy 364 KPHLLQDA--DKQFOOLOQAKVAMEIPAM-----LYDTRNNWEIDRGLERG-- 409
Db 292 KEQYRLDEALPDHETVDAERARLAGDRDAMRSVVGALCEELVRIKDSLDLFRVSRGHP 351
Qy 410 --LCALL--1GKVDCEKRWLGDSQYRNPAIVEFVLNSNRDND-----DLPGLCKL 461
Db 352 SELDALLAPLKQIADTLAVLFGCPQPKVILDQDVIHALAQRREBPSDAILMDVAGALLY 411
Qy 462 LETWLAGVVFP--RFRDTKDKKFKLGDYD-----DPMVLSYL----- 497
Db 412 VEATLAGMAGPGDERSERLPTTDVAQIHLVIKEARNGLEQAKDAIIEFTASQWNHE 471
Qy 498 -----ERVEVQGS-----PLAAATMARIAGHVKASAMQALQKVP----- 535
Db 472 HLARVPELLTQVRGGLAMIPLERAATLLETNCNRYIQEQLL--ARKAVPDWQSLDTLADAI 529
Qy 536 ---SRVTDNSABPKDVQETVFSV--DPVGN-----NVGRDGEPCGVFTAEAVRPSNP 583
Db 530 TSVEYLLERUSEDHASQSDLLDVAEDSLANGLYTLKPNSSAPAEPL----- 577
Qy 584 ETNDVAIRAGVSSSDVETTVMSVADMLKEASVKILAAGVAIGLISLFSQYFLKSSSS 643
Db 578 -SGPAALSPAAEPERPEAVVE--VAETAQPPADTAPAEA----- 616
Qy 644 FORKDWSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEML 703
Db 617 --REDAPQLASDNNWTLGEVAPDAGE--PSLDL-----ALDPLDSDSAEVP 658
Qy 704 PEVLDCRMLKIWDRAEAETLAQLGVVDYTLKLSVDSVTVSAD-----GTRALVEATLER 758
Db 659 PAVPECRE-----TPPOSTSAPARSLDLSLDEIDLSGLDLFPADAPASGPAALADWSLPE 714
Qy 759 SACL--SDLVHP 768
Db 715 OWGLGDDLAQP 725
RESULT 13
S70368
DNA polymerase I - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70368
R;Riggs, M.G.; Tudor, S.; Sivaram, M.; McDonough, S.H.
Biochim. Biophys. Acta 1307, 178-186, 1996
A;Title: Construction of single amino acid substitution mutants of cloned Bacillus : team
A;Reference number: S70368; MUID:96283813; PMID:8679703
A;Accession: S70368
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-876 <RIG>
A;Cross-references: UNIPROT:P52026; UNIPARC:UPI000016E912; EMBL:L42111; NID:9806280; PID
C;Genetics:
A;Start codon: TTG
C;Superfamily: DNA-directed DNA polymerase I
Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 3.6;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
Qy 99 HFLTDGIRAFARVSKPPQFGFSDDALISRRQILQAACETLNSRPSRREY-NEGLLDDE 157
Db 107 HYEADDICTMAAREE---GFAVKVISGDRDLTQ-----LASPQVVEITKKGITDIE 158
Qy 158 EATVITDV-----PMDKVPF-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTPTTVEKYGLTPEQIVDLKGLMGKSDNIPGVPGIGEKTAVKLLKQFGTVENVLASI 218

Qy 191 EALLKRLPKSFQ--DVLVLMALAFDVSRLA---MALDPPDPFITGYFVEEALKLQE 245
Db 219 DEIKGEKLENLRQYRDAL--LSQLAAICRDAPEVELTLD--DIVYKGEDREKVVALLQE 275
Qy 246 EGASSLAPDLRAQIDE-----TLBEIIPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFDRKMAVQTDGEKPLAGMDFAIADSVTDEMLADKAALVVVGVGDYNNHIAPIVGI 335
Qy 291 -----LSGVNRLWSVGGGASALVGGLTREKPMNEAFLKMTAAE----- 330
Db 336 ALANERGRFPLRPETALADPKFLAM-----LGDETCKKTMFDSKRAVALKWKGI 385
Qy 331 -----QVDLFVATPSNIPAES-----FEYVEVALALVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVVFLLLAAYLLDPAQAAGVAAKMHQVEAVRSDEAVYKGAKRTPVDBPTLAE 445
Qy 371 DADQKQFOOLOQAKVAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDCRM 423
Db 446 HLAR-----KAAAIWALEBPLMDLRRNEQDRLLTLELEQPLAGILANMEFTGVKVDTKRL 500
Qy 424 -WGLD--SEDSQYRNPAIVEFVLNSNRDNDLPLGLCKLETLWLAGVVFPRFRDTKDKK 481
Db 501 EQMGAELTEQLQ-----AVERRIYELAQEENINSP---KQLGT---VLFDKLQLPVLKK 549
Qy 482 FKLGDIYDDPMVLSYL--ERVEVQGSPLAAAAATWARIAGHVKASAMQALQKVPFSRYTD 540
Db 550 TKTG-YSTSADVLEKLAPHHEIVE--HILHYRQLGKLQSTVIE--GLLKVVHVPVTKVHTM 605
Qy 541 RNSAEPKDVQET--VFSVDVPGNVGDRDGEFVPIAEAVRPE-----NFEINDVAI 590
Db 606 FNOA-----LTQGLSSVEPNLQNIPIRLBGRKIROAFVFPSEPDWLIFFAADYSQIELRV 661
Qy 591 RAGVSSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLFSQ 634
Db 662 LAHAE---DDNLEAPRRGLDHTKTAMDI FHVSEEDVTANMRROAKAVNFGVYIGSD 718
Qy 635 KYFLKSSSSQFRKDWSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQLNLTIRKEAAEFIERYPASFPQVK-----QYMDNIVQE-AKQKGYVT 764
Qy 695 GPDHRIEMLPEVLDCRM-LKIWDTRAA-ETAQLGLVVDYTLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRRYLPDITSRNFNVRSAERTAMNTPIQSAAD-IKKAMIDLSVRLREERLOAR 823
Qy 751 -----LVEATLRESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 14

E83594

still frameshift probable component of chemotactic signal transduction system PA0413 [im

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C;Accession: E83594

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: E83594

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2472 <STO>

A;Cross-references: UNIPROT:Q91696; UNIPARC:UPI00000C5045; GB:AE004478; GB:AE004091; NID:

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0413

Query Match 3.1%; Score 127.5; DB 2; Length 2472;

Best Local Similarity 20.0%; Pred. No. 22;

Matches	159;	Conservative	105;	Mismatches	291;	Indels	239;	Gaps	36;
Qy	124	DALISRRQILQAACETL	SNPRSRREYN	EGLLDD	BEATVITDP	WDKVP	GCALCVL	QEGGE-	182
Db	19	ETLKQAKALEAF	VENPQDP	TRMR	-----	FCLTVV--	HQVQ	GLQWVEFYGAA	64
Qy	183	--TEIVL	RVG	EALLKERL	PKSPKQDV	LVLMALAF	LDVSR	DAMALDP	PPDF
Db	65	LIABEM	QVALLDGR	VPNQGEAL	EVLMQAILQ	PVYLDRIQ	TARRDL	PMVVLPL	DLNDL
Qy	235	FVEERALK	LLQEG	ASSLAP	LD--	RAQID-	-----	ETLEEITPR	VLELLGLP
Db	125	RAARGEK	LLSE--	TSLFAP	DL	SQRPQLD	GEAIAQ	LRTDEL	GGLLRKL
Qy	279	LGDDY	YAAK----	RINGL	SGVRNI--	LMSVGGG	GAASALV	CGGLTRE	KPMERAF
Db	183	RNQDVAT	SLGYLAR	VYARLE	GLCREA	PLGLWSI	-----	ASGLVE	GLANGSVVNSASVR
Qy	328	AAEQVD-	-----	LFVAT	PSNI-	-----	PAESPE	VYEVALAL	VAQAFIG
Db	238	LLRQLD	RELKEL	VSQAG	DLNQAP	DELVK	ULLFY	VAK	SPQSPIRAL
Qy	364	KXPHLL	QDA--	DKQFQ	Q	QAKV	MAMEIP	AM-----	LYDTRNN
Db	287	KEQYRL	DEALPD	HETVDA	ERAAR	L	AGPDR	AMRSV	VVGALCEEL
Qy	410	--LCALL--	ICKVDE	CRMW	LGLD	SEDSQ	YRNP	PAIVEF	VLENSNR
Db	347	SEL	DALLAP	LKQI	ADTTLAV	LGFG	QPKVIL	QDLDV	IHALA
Qy	462	LETWLAG	VVFP--	RPRDT	KDKK	FKLGD	VYD-----	DPMW	LSYL-----
Db	407	VEATLAG	MAGG	CDER	ENSE	SRLPTD	VQA	IHQLV	IKEARNG
Qy	498	-----	BRVEV	QGS-----	PLAA	AATW	ARIGAE	HVKAS	AMQALQKVP
Db	467	HLARV	PELLTQ	VRGGL	AMIPLER	AATLT	ETCN	RYIQ	EQLL--
Qy	536	--SRYT	DRNSA	AEPKD	VOET	VFSV--	DPVGN-	-----	NVGRCE
Db	525	TSVEY	YLER	LSE	HD	SHASG	SDIL	VAED	SLANL
Qy	584	ETNDY	AIRAG	VSESS	VDTE	TVE	MSVAD	MLKEAS	VKILAAG
Db	573	--SGPAA	IESPA	AEPE	PEAVVE--	VAETA	SQPPAD	TAPAEAA	-----
Qy	644	FOR	KDWM	S	MES	VDAT	TGSR	AD	SEALPRMD
Db	612	--REDAP	QLAS	DDN	WTG	EVAP	DAGE--	PSL	DL-----
Qy	703	--LPEV	LDGR	MLK	KTWD	RAA	ETAQ	GLV	YDYV
Db	654	PALPE	VVEES	GQ	QS	T	SAP	ARS	LD-----
Qy	756	LEES	ACL	-	SDL	VHP	768		
Db	709	LPEQ	WGLG	DD	LAQ	P	722		

```

RESULT 15
T30853
antigenic heat-stable protein - Rickettsia japonica
C/Species: Rickettsia japonica
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C/Accession: T30853
R/Uchiyama, T.
submitted to the EMBL Data Library, May 1997
A/Description: Sequencing of the gene encoding the antigenic heat-stable 120-kilodalton
A/Reference number: Z20905
A/Accession: T30853
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1018 <UCH>

```

A:Cross-references: UNIPROT:Q9AJ79; UNIPARC:UPI0000135607; EMBL:AB003696; NID:d11066377;

Query Match	3.1%	Score 126.5;	DB 2;	Length 1018;
Best Local Similarity	19.2%	Pred. No. 6.2;		
Matches 128;	Conservative 109;	Mismatches 265;	Indels 163;	Gaps 28;

QY	115	KPQPGFSDALISRRQILQAACETLNP	RRREYN	EGLLDDEEATVITDVPWDKVP	PGAL 174
Db	374	EPPOYKQVPPITPTNPLOP--ETSOM	POS-QQVNP	LLN--AATALSGSQD	LLNYN 428
QY	175	CVLQEGGETEIVLRVGEALIKERL	PKSFQD	VVLVMA--LAPLDYSRD	AMALDPPDPFITG 232
Db	429	AGLTK--EIDLKEA	TAIILNDK-----KSDIAEKQ	ANIILAENTVNNK	ITPDAKVAG 481
QY	233	YEFVEEALKLQEGGASSLAPDLR--	-----AQIDETLEEITPRYVLEL--	-----LGLP	278
Db	482	VNAVLETIKNDQN-----TPDLEKS	KMLEATVAIALNSENLEPKQ	KQKQILKSAVDGLS	535
QY	279	LGDDYAAKELNGLSGVNRNII	LVGGGASALVGLTREKFWNEAFLR--	MTAAEQVD	LFV 336
Db	536	LKDD-----ASRAA	AIDGIT-----DAVIKNSL	STEDKGTMI	568
QY	337	ATPSNIPAESFEYVEVALVAQAF	IGKPHLLQDADKQFQOLQQA	KVMA-----ME	388
Db	569	AVGDKVNVSELSNAEKQLGSLV	KKGVEAQVLSPAQQQLMQONL	DKITAEQTKKDTIKK	628
QY	389	IPAMLYDTRNWE-----IDFGL	ERGLCALLIGKVDECRMWLG	LSDESDSQYR	435
Db	629	VNDILFDPLSNLTSLKNTIQAII	SNVLDDGPATAEVKGEIIQEIT	NTVAGSSLEAHD----	684
QY	436	NPAIVEFVLEN--SNRDDNDL	PLGCKLLETWLAGVVPFRD	TKDKKFKGLGDYDDPMV	493
Db	685	KAAIINGISETIATHSDTSLSP	NKALIMASAEKGA-----ESQ	NLPD--RELMT	734
QY	494	LSYLERVEVQGSPLAAAAATMAR	IGAHHVKASAMQALQK-----	VFPSRYTDRNSA-----	544
Db	735	KGLVDGIVEGKGPEITTKAVSS	GIDNSINIDSEKEALKKAKDA	ASEAALDRETQNLTEGL	794
QY	545	-----EPKD-----VOET	VSPDVPGVNNGRDP	GEVFAE-----	575
Db	795	KQNTIEHKPHDDIYNKREVINA	NPVIBALEKSEKFPVW	SABERIVQETSSILN	ISKL 854
QY	576	AVRPSENFETNDYAIRAGVS	ESVDTEVTMSVADMLKEASV	KILAAGVAIGLISL	FSQK 635
Db	855	AVEKYNPF-----RAMLS	SPNGNLTKLEK-----KEESI	KKVDE-----LVKAF	GTGK 896
QY	636	YFLKSSSSSFQKDMVS-----	SMESDVATIGSV	RADD--SEALPRMD	ARTAE-NIVSKWKQK 688
Db	897	SSTEQSQFIKANLDDKTLSE	IRLQTIKDLQEKRAEAIENP	SVKTEDVRVVS	GSKS 956
QY	689	IKSLA 693			
Db	957	LKPI5 961			

Search completed: December 4, 2005, 12:50:55
Job time : 49 secs

Search completed: December 4, 2005, 12:50:55
Job time : 49 secs